

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

us
C31

~~CHLAMYDIA PNEUMONIAE GENOME SEQUENCE~~

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application is related to 60/128,606, filed April 8, 1999 and
5 60/108,279, filed November 12, 1998, which are incorporated herein by reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

10 FIELD OF THE INVENTION

This invention relates to nucleic acids and polypeptides from *Chlamydia pneumoniae* and to their use in the diagnosis, prevention and treatment of diseases associated with *C. pneumoniae*.

15 BACKGROUND OF THE INVENTION

Chlamydiaceae is a family of obligate intracellular parasite with a tropism for epithelial cells lining the mucus membranes. The bacteria have two morphologically distinct forms, "elementary body" and "reticulate body". The elementary body is the infectious form, and has a rigid cell wall, primarily of cross-linked outer membrane
20 proteins. The reticulate body is the intracellular, metabolically active form. A unique developmental cycle between these two forms characterizes *Chlamydia* growth.

C. pneumoniae is a human respiratory pathogen that causes acute respiratory disease, and approximately 10% of community-acquired pneumonia. Antibody prevalence studies have shown that virtually everyone is infected with *C.*
25 *pneumoniae* at some time, and that reinfection is common. In addition to respiratory disease, studies have shown an association of this organism with coronary artery disease. It has been demonstrated in atherosclerotic lesions of the aorta and coronary arteries by immunocytochemistry and by polymerase chain reaction (Kuo *et al.* (1993) *J Infect Dis* 167(4):841-849).

Recent reports have further demonstrated the presence of *C. pneumoniae* in the walls of abdominal aortic aneurysms (Juvonen *et al.* (1997) *J Vasc Surg*
30 25(3):499-505). Abdominal aortic aneurysms are frequently associated with atherosclerosis, and inflammation may be an important factor in aneurysmal dilatation.

C. pneumoniae may play a role in maintaining an inflammation and triggering the development of aortic aneurysms.

Muhlestein *et al.* (1996) JACC 27:1555-61, reported a differential incidence of *Chlamydia* species within the coronary artery wall of patients with atherosclerosis versus those with other forms of cardiovascular disease. The extremely high rate of possible infection in patients with symptomatic atherosclerotic disease compared to the very low rate in patients with normal coronary arteries or coronary artery disease from chronic transplant rejection provides evidence for a direct link between the atherosclerotic process and *Chlamydia* infection. Because a history of chlamydial infection is so prevalent in the population, the issue of causality remains. On a physiologic and pathologic level, abnormal interactions among endothelial cells, platelets, macrophages and lymphocytes may lead to a cascade of events resulting in acute endothelial damage, thrombosis and repair, chronically leading to the development of atheroma in blood vessels.

C. pneumoniae is related to other *Chlamydia* species, but the level of sequence similarity is relatively low. Very little is known about the biology of this organism, although it appears to be an important human pathogen. Allelic diversity and structural relationships between specific genes of Chlamydial species is described in Kaltenboeck *et al.* (1993) J Bacteriol 175(2):487-502; Gaydos *et al.* (1992) Infect Immun 60(12):5319-5323; Everett *et al.* (1997) Int J Syst Bacteriol 47(2):461-473; and Pudjiatmoko *et al.* (1997) Int J Syst Bacteriol 47(2):425-431.

A number of studies have been published describing methods for detection of *C. pneumoniae*, and for distinguishing between Chlamydial species. Such methods include PCR detection (Rasmussen *et al.* (1992) Mol Cell Probes 6(5):389-394; Holland *et al.* (1990) J Infect Dis 162(4):984-987); a simplified polymerase chain reaction-enzyme immunoassay (Wilson *et al.* (1996) J Appl Bacteriol 80(4):431-438); sequence determination and restriction endonuclease cleavage (Herrmann *et al.* (1996) J Clin Microbiol 34(8):1897-1902).

Antigenic and molecular analyses of different *C. pneumoniae* strains is described in Jantos *et al.* (1997) J Clin Microbiol 35(3):620-623. Some genes of *C. pneumoniae* have been isolated and sequenced. These include the Gro E operon (Kikuta *et al.* (1991) Infect Immun 59(12):4665-4669); the major outer membrane protein Perez *et*

al. (1991) Infect Immun 59(6):2195-2199; the DnaK protein homolog (Kornak *et al.* (1991) Infect Immun 59(2):721-725); as well as a number of ribosomal and other genes.

5

SUMMARY OF THE INVENTION

This invention provides the genomic sequence of *Chlamydia pneumoniae*.

The sequence information is useful for a variety of diagnostic and analytical methods.

The genomic sequence may be embodied in a variety of media, including computer
10 readable forms, or as a nucleic acid comprising a selected fragment of the sequence.
Such fragments generally consist of an open reading frame, transcriptional or translational control elements, or fragments derived therefrom. Proteins encoded by the open reading frames are useful for diagnostic purposes, as well as for their enzymatic or structural activity.

15

DEFINITIONS

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those
20 encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group., e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl
25 sulfonium Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

30

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Amplification" primers are oligonucleotides comprising either natural or analogue nucleotides that can serve as the basis for the amplification of a select nucleic acid sequence. They include, e.g., polymerase chain reaction primers and ligase chain reaction oligonucleotides.

5 "Antibody" refers to an immunoglobulin molecule able to bind to a specific epitope on an antigen. Antibodies can be a polyclonal mixture or monoclonal. Antibodies can be intact immunoglobulins derived from natural sources or from recombinant sources and can be immunoreactive portions of intact immunoglobulins. Antibodies may exist in a variety of forms including, for example, Fv, F_{ab}, and F(ab)₂, as
10 well as in single chains. Single-chain antibodies, in which genes for a heavy chain and a light chain are combined into a single coding sequence, may also be used.

An "antigen" is a molecule that is recognized and bound by an antibody, e.g., peptides, carbohydrates, organic molecules, or more complex molecules such as glycolipids and glycoproteins. The part of the antigen that is the target of antibody
15 binding is an antigenic determinant and a small functional group that corresponds to a single antigenic determinant is called a hapten.

"Biological sample" refers to any sample obtained from a living or dead organism. Examples of biological samples include biological fluids and tissue specimens. Such biological samples can be prepared for analysis of the presence of *C. pneumoniae*
20 nucleic acids, proteins, or antibodies specifically reactive with the proteins.

The term "*C. pneumoniae* gene" shall be intended to mean the open reading frame encoding specific *C. pneumoniae* polypeptides, as well as adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 2 kb beyond the coding region, but possibly further in either direction. The gene may be
25 introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially
30 identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues

(Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid.

Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 2) Serine (S), Threonine (T);
- 3) Aspartic acid (D), Glutamic acid (E);
- 4) Asparagine (N), Glutamine (Q);
- 5) Cysteine (C), Methionine (M);
- 6) Arginine (R), Lysine (K), Histidine (H);
- 7) Isoleucine (I), Leucine (L), Valine (V); and
- 8) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

see, e.g., Creighton, *Proteins* (1984)).

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a
 5 comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. This definition also refers to the complement of a test sequence, which has a designated percent sequence or subsequence complementarity when the test sequence has a designated or substantial identity to a reference sequence. For example, a designated amino acid percent identity
 10 of 95% refers to sequences or subsequences that have at least about 95% amino acid identity when aligned for maximum correspondence over a comparison window as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences would then be said to have substantial identity, or to be substantially identical to each other. Preferably, sequences have at least
 15 about 70% identity, more preferably 80% identity, more preferably 90-95% identity and above. Preferably, the percent identity exists over a region of the sequence that is at least about 25 amino acids in length, more preferably over a region that is 50-100 amino acids in length.

When percentage of sequence identity is used in reference to proteins or
 20 peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may
 25 be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a
 30 conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA)..

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated or default program parameters.

10 A comparison window includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 25 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison
15 can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics
20 Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Ausubel et al., supra*).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or
25 dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment
30 procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The

final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, e.g, version 7.0 (Devereaux *et al.*, *Nuc. Acids Res.* 12:387-395 (1984).

Another example of algorithm that is suitable for determining percent sequence identity (i.e., substantial similarity or identity) is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues, always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as default parameters a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

Another indication that polynucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically stringent conditions for a Southern blot protocol involve hybridizing in a buffer comprising 5x SSC, 1% SDS at 65°C or hybridizing in a buffer containing 5x SSC and 1% SDS at 42°C and washing at 65°C with a 0.2x SSC, 0.1% SDS wash.

A "label" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ³²P, fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, dioxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available.

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

As used herein a "nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (i.e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are preferably directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence.

A labeled nucleic acid probe or oligonucleotide is one that is bound, either covalently, through a linker, or through ionic, van der Waals or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

"Pharmaceutically acceptable" means a material that is not biologically or otherwise undesirable, i.e., the material can be administered to an individual along with a *Chlamydia* antigen without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical composition.

5 The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers.

10 The phrase "specifically or selectively hybridizing to," refers to hybridization between a probe and a target sequence in which the probe binds substantially only to the target sequence, forming a hybridization complex, when the target is in a heterogeneous mixture of polynucleotides and other compounds. Such hybridization is determinative of the presence of the target sequence. Although the probe
15 may bind other unrelated sequences, at least 90%, preferably 95% or more of the hybridization complexes formed are with the target sequence.

The term "recombinant" when used with reference to a cell, or nucleic acid, or vector, indicates that the cell, or nucleic acid, or vector, has been modified by the introduction of a heterologous nucleic acid or the alteration of a native nucleic acid, or
20 that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

The phrase "specifically immunoreactive with", when referring to a protein
25 or peptide, refers to a binding reaction between the protein and an antibody which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other compounds. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such
30 conditions may require an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein and are described in detail below.

The phrase "substantially pure" or "isolated" when referring to a *Chlamydia* peptide or protein, means a chemical composition which is free of other subcellular components of the *Chlamydia* organism. Typically, a monomeric protein is substantially pure when at least about 85% or more of a sample exhibits a single polypeptide backbone. Minor variants or chemical modifications may typically share the same polypeptide sequence. Depending on the purification procedure, purities of 85%, and preferably over 95% pure are possible. Protein purity or homogeneity may be indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band on a polyacrylamide gel upon silver staining. For certain purposes high resolution will be needed and HPLC or a similar means for purification utilized.

DETAILED DESCRIPTION

The present invention provides the nucleotide sequence of the *C. pneumoniae* genome SEQ ID NO: 1 or a representative fragment thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. As used herein, a "representative fragment" of the nucleotide sequence depicted in SEQ ID NO: 1 refers to any portion which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are open reading frames, expression modulating fragments, uptake modulating fragments, and fragments which can be used to diagnose the presence of *C. pneumoniae* in sample. Using the information provided in the present application, together with routine cloning and sequencing methods, one of ordinary skill in the art will be able to clone and sequence all "representative fragments" of interest including open reading frames (ORFs) encoding a large variety of *C. pneumoniae* proteins. A non-limiting identification of such preferred representative fragments is provided in Tables 2 and 3.

Diagnostic use of *C. pneumoniae* nucleic acids

Hybridization-based assays

Using the nucleic acids disclosed here, one of skill can design nucleic acid hybridization-based assays for the detection of *C. pneumoniae*. Any of a number of well known techniques for the specific detection of target nucleic acids can be used. Exemplary hybridization-based assays include, but are not limited to, traditional "direct

probe" methods such as Southern Blots, dot blots, *in situ* hybridization (e.g., FISH), PCR, and the like. The methods can be used in a wide variety of formats including, but not limited to substrate- (e.g. membrane or glass) bound methods or array-based approaches as described below. As noted above, this invention also embraces methods for detecting the presence of *Chlamydia* DNA or RNA in biological samples. These sequences can be used to detect *Chlamydia* in biological samples from patients suspected of being infected. A variety of methods of specific DNA and RNA measurement using nucleic acid hybridization techniques are known to those of skill in the art (see Sambrook *et al.*, *supra*).

In situ hybridization assays are well known (e.g., Angerer (1987) *Meth. Enzymol* 152: 649). Generally, *in situ* hybridization comprises the following major steps: (1) fixation of tissue or biological structure to be analyzed; (2) prehybridization treatment of the biological structure to increase accessibility of target DNA, and to reduce nonspecific binding; (3) hybridization of the mixture of nucleic acids to the nucleic acid in the biological structure or tissue; (4) post-hybridization washes to remove nucleic acid fragments not bound in the hybridization and (5) detection of the hybridized nucleic acid fragments. The reagent used in each of these steps and the conditions for use vary depending on the particular application.

In a typical *in situ* hybridization assay, cells are fixed to a solid support, typically a glass slide. If a nucleic acid is to be probed, the cells are typically denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of labeled probes specific to the nucleic acid sequence encoding the protein. The targets (e.g., cells) are then typically washed at a predetermined stringency or at an increasing stringency until an appropriate signal to noise ratio is obtained.

The nucleic acids of this invention are particularly well suited to array-based hybridization formats. Arrays are a multiplicity of different "probe" or "target" nucleic acids (or other compounds) attached to one or more surfaces (e.g., solid, membrane, or gel). In a preferred embodiment, the multiplicity of nucleic acids (or other moieties) is attached to a single contiguous surface or to a multiplicity of surfaces juxtaposed to each other.

In an array format a large number of different hybridization reactions can be run essentially "in parallel." This provides rapid, essentially simultaneous, evaluation

of a number of hybridizations in a single "experiment". Methods of performing hybridization reactions in array based formats are well known to those of skill in the art (see, e.g., Pastinen (1997) *Genome Res.* 7: 606-614; Jackson (1996) *Nature Biotechnology* 14:1685; Chee (1995) *Science* 274: 610; WO 96/17958.

5 Arrays, particularly nucleic acid arrays can be produced according to a wide variety of methods well known to those of skill in the art. For example, in a simple embodiment, "low density" arrays can simply be produced by spotting (e.g. by hand using a pipette) different nucleic acids at different locations on a solid support (e.g. a glass surface, a membrane, etc.).

10 This simple spotting, approach has been automated to produce high density spotted arrays (see, e.g., U.S. Patent No: 5,807,522). This patent describes the use of an automated systems that taps a microcapillary against a surface to deposit a small volume of a biological sample. The process is repeated to generate high density arrays. Arrays can also be produced using oligonucleotide synthesis technology. Thus, for
15 example, U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092 teach the use of light-directed combinatorial synthesis of high density oligonucleotide arrays.

20 Many methods for immobilizing nucleic acids on a variety of solid surfaces are known in the art. A wide variety of organic and inorganic polymers, as well as other materials, both natural and synthetic, can be employed as the material for the solid surface. Illustrative solid surfaces include, e.g., nitrocellulose, nylon, glass, quartz, diazotized membranes (paper or nylon), silicones, polyformaldehyde, cellulose, and cellulose acetate. In addition, plastics such as polyethylene, polypropylene, polystyrene, and the like can be used. Other materials which may be employed include paper,
25 ceramics, metals, metalloids, semiconductive materials, cermets or the like. In addition, substances that form gels can be used. Such materials include, e.g., proteins (e.g., gelatins), lipopolysaccharides, silicates, agarose and polyacrylamides. Where the solid surface is porous, various pore sizes may be employed depending upon the nature of the system.

30 In preparing the surface, a plurality of different materials may be employed, particularly as laminates, to obtain various properties. For example, proteins (e.g., bovine serum albumin) or mixtures of macromolecules (e.g., Denhardt's solution) can be employed to avoid non-specific binding, simplify covalent conjugation, enhance

signal detection or the like. If covalent bonding between a compound and the surface is desired, the surface will usually be polyfunctional or be capable of being polyfunctionalized. Functional groups which may be present on the surface and used for linking can include carboxylic acids, aldehydes, amino groups, cyano groups, ethylenic groups, hydroxyl groups, mercapto groups and the like. The manner of linking a wide variety of compounds to various surfaces is well known and is amply illustrated in the literature.

For example, methods for immobilizing nucleic acids by introduction of various functional groups to the molecules is known (see, e.g., Bischoff (1987) *Anal. Biochem.*, 164: 336-344; Kremsky (1987) *Nucl. Acids Res.* 15: 2891-2910). Modified nucleotides can be placed on the target using PCR primers containing the modified nucleotide, or by enzymatic end labeling with modified nucleotides. Use of glass or membrane supports (e.g., nitrocellulose, nylon, polypropylene) for the nucleic acid arrays of the invention is advantageous because of well developed technology employing manual and robotic methods of arraying targets at relatively high element densities. Such membranes are generally available and protocols and equipment for hybridization to membranes is well known.

Target elements of various sizes, ranging from 1 mm diameter down to 1 μm can be used. Smaller target elements containing low amounts of concentrated, fixed probe DNA are used for high complexity comparative hybridizations since the total amount of sample available for binding to each target element will be limited. Thus it is advantageous to have small array target elements that contain a small amount of concentrated probe DNA so that the signal that is obtained is highly localized and bright. Such small array target elements are typically used in arrays with densities greater than $10^4/\text{cm}^2$. Relatively simple approaches capable of quantitative fluorescent imaging of 1 cm^2 areas have been described that permit acquisition of data from a large number of target elements in a single image (see, e.g., Wittrup (1994) *Cytometry* 16:206-213).

If fluorescently labeled nucleic acid samples are used, arrays on solid surface substrates with much lower fluorescence than membranes, such as glass, quartz, or small beads, can achieve much better sensitivity. Substrates such as glass or fused silica are advantageous in that they provide a very low fluorescence substrate, and a highly efficient hybridization environment. Covalent attachment of the target nucleic acids to glass or synthetic fused silica can be accomplished according to a number of

known techniques (described above). Nucleic acids can be conveniently coupled to glass using commercially available reagents. For instance, materials for preparation of silanized glass with a number of functional groups are commercially available or can be prepared using standard techniques (see, *e.g.*, Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Wash., D.C.). Quartz cover slips, which have at least 10-fold lower autofluorescence than glass, can also be silanized.

Alternatively, probes can also be immobilized on commercially available coated beads or other surfaces. For instance, biotin end-labeled nucleic acids can be bound to commercially available avidin-coated beads. Streptavidin or anti-digoxigenin antibody can also be attached to silanized glass slides by protein-mediated coupling using *e.g.*, protein A following standard protocols (see, *e.g.*, Smith (1992) *Science* 258: 1122-1126). Biotin or digoxigenin end-labeled nucleic acids can be prepared according to standard techniques. Hybridization to nucleic acids attached to beads is accomplished by suspending them in the hybridization mix, and then depositing them on the glass substrate for analysis after washing. Alternatively, paramagnetic particles, such as ferric oxide particles, with or without avidin coating, can be used.

A variety of other nucleic acid hybridization formats are known to those skilled in the art. For example, common formats include sandwich assays and competition or displacement assays. Hybridization techniques are generally described in Hames and Higgins (1985) *Nucleic Acid Hybridization, A Practical Approach*, IRL Press; Gall and Pardue (1969) *Proc. Natl. Acad. Sci. USA* 63: 378-383; and John *et al.* (1969) *Nature* 223: 582-587.

Sandwich assays are commercially useful hybridization assays for detecting or isolating nucleic acid sequences. Such assays utilize a "capture" nucleic acid covalently immobilized to a solid support and a labeled "signal" nucleic acid in solution. The sample will provide the target nucleic acid. The "capture" nucleic acid and "signal" nucleic acid probe hybridize with the target nucleic acid to form a "sandwich" hybridization complex. To be most effective, the signal nucleic acid should not hybridize with the capture nucleic acid.

Detection of a hybridization complex may require the binding of a signal generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a ligand-conjugated probe and an anti-ligand conjugated with a signal.

04438435-4449
06777-5272760

The sensitivity of the hybridization assays may be enhanced through use of a nucleic acid amplification system that multiplies the target nucleic acid being detected. Examples of such systems include the polymerase chain reaction (PCR) system and the ligase chain reaction (LCR) system. Other methods recently described in the art are the nucleic acid sequence based amplification (NASBAO, Cangene, Mississauga, Ontario) and Q Beta Replicase systems.

Nucleic acid hybridization simply involves providing a denatured probe and target nucleic acid under conditions where the probe and its complementary target can form stable hybrid duplexes through complementary base pairing. The nucleic acids that do not form hybrid duplexes are then washed away leaving the hybridized nucleic acids to be detected, typically through detection of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids, or in the addition of chemical agents, or the raising of the pH. Under low stringency conditions (e.g., low temperature and/or high salt and/or high target concentration) hybrid duplexes (e.g., DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (e.g., higher temperature or lower salt) successful hybridization requires fewer mismatches.

One of skill in the art will appreciate that hybridization conditions may be selected to provide any degree of stringency. In a preferred embodiment, hybridization is performed at low stringency to ensure hybridization and then subsequent washes are performed at higher stringency to eliminate mismatched hybrid duplexes. Successive washes may be performed at increasingly higher stringency (e.g., down to as low as 0.25 X SSPE-T at 37°C to 70°C) until a desired level of hybridization specificity is obtained. Stringency can also be increased by addition of agents such as formamide. Hybridization specificity may be evaluated by comparison of hybridization to the test probes with hybridization to the various controls that can be present.

In general, there is a tradeoff between hybridization specificity (stringency) and signal intensity. Thus, in a preferred embodiment, the wash is performed at the highest stringency that produces consistent results and that provides a signal intensity greater than approximately 10% of the background intensity. Thus, in a preferred embodiment, the hybridized array may be washed at successively higher

stringency solutions and read between each wash. Analysis of the data sets thus produced will reveal a wash stringency above which the hybridization pattern is not appreciably altered and which provides adequate signal for the particular probes of interest.

Methods of optimizing hybridization conditions are well known to those of skill in the art (see, e.g., Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes*, Elsevier, N.Y.).

Labeling and detection of nucleic acids.

In a preferred embodiment, the hybridized nucleic acids are detected by detecting one or more labels attached to the sample or probe nucleic acids. The labels may be incorporated by any of a number of means well known to those of skill in the art. Means of attaching labels to nucleic acids include, for example nick translation or end-labeling (e.g. with a labeled RNA) by kinasing of the nucleic acid and subsequent attachment (ligation) of a nucleic acid linker joining the sample nucleic acid to a label (e.g., a fluorophore). A wide variety of linkers for the attachment of labels to nucleic acids are also known. In addition, intercalating dyes and fluorescent nucleotides can also be used.

Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., DynabeadsTM), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like, see, e.g., Molecular Probes, Eugene, Oregon, USA), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold (e.g., gold particles in the 40 -80 nm diameter size range scatter green light with high efficiency) or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

A fluorescent label is preferred because it provides a very strong signal with low background. It is also optically detectable at high resolution and sensitivity through a quick scanning procedure. The nucleic acid samples can all be labeled with a single label, e.g., a single fluorescent label. Alternatively, in another embodiment, different nucleic acid samples can be simultaneously hybridized where each nucleic acid

target nucleic acid prior to the hybridization. Thus, for example, the target nucleic acid may be biotinylated before the hybridization. After hybridization, an avidin-conjugated fluorophore will bind the biotin bearing hybrid duplexes providing a label that is easily detected. For a detailed review of methods of labeling nucleic acids and detecting labeled
 5 hybridized nucleic acids see *Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes*, P. Tijssen, ed. Elsevier, N.Y., (1993)).

Fluorescent labels are easily added during an *in vitro* transcription reaction. Thus, for example, fluorescein labeled UTP and CTP can be incorporated into
 10 the RNA produced in an *in vitro* transcription.

The labels can be attached directly or through a linker moiety. In general, the site of label or linker-label attachment is not limited to any specific position. For example, a label may be attached to a nucleoside, nucleotide, or analogue thereof at any position that does not interfere with detection or hybridization as desired. For example,
 15 certain Label-ON Reagents from Clontech (Palo Alto, CA) provide for labeling interspersed throughout the phosphate backbone of an oligonucleotide and for terminal labeling at the 3' and 5' ends. As shown for example herein, labels can be attached at positions on the ribose ring or the ribose can be modified and even eliminated as desired. The base moieties of useful labeling reagents can include those that are naturally
 20 occurring or modified in a manner that does not interfere with the purpose to which they are put. Modified bases include but are not limited to 7-deaza A and G, 7-deaza-8-aza A and G, and other heterocyclic moieties.

It will be recognized that fluorescent labels are not to be limited to single species organic molecules, but include inorganic molecules, multi-molecular mixtures of
 25 organic and/or inorganic molecules, crystals, heteropolymers, and the like. Thus, for example, CdSe-CdS core-shell nanocrystals enclosed in a silica shell can be easily derivatized for coupling to a biological molecule (Bruchez *et al.* (1998) *Science*, 281: 2013-2016). Similarly, highly fluorescent quantum dots (zinc sulfide-capped cadmium selenide) have been covalently coupled to biomolecules for use in ultrasensitive
 30 biological detection (Warren and Nie (1998) *Science*, 281: 2016-2018).

Amplification-based assays.

In another embodiment, amplification-based assays can be used to detect nucleic acids. In such amplification-based assays, the nucleic acid sequences act as a

template in an amplification reaction (e.g. Polymerase Chain Reaction (PCR). Detailed protocols for quantitative PCR are provided in Innis *et al.* (1990) *PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc. N.Y.).

Other suitable amplification methods include, but are not limited to ligase chain reaction (LCR) (see Wu and Wallace (1989) *Genomics* 4: 560, Landegren *et al.* (1988) *Science* 241: 1077, and Barringer *et al.* (1990) *Gene* 89: 117, transcription amplification (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173), and self-sustained sequence replication (Guatelli *et al.* (1990) *Proc. Nat. Acad. Sci. USA* 87: 1874).

Detection of *C. pneumoniae* gene expression

The nucleic acids of the invention can also be used to *C. pneumoniae* detect gene transcripts. Methods of detecting and/or quantifying gene transcripts using nucleic acid hybridization techniques are known to those of skill in the art (see Sambrook *et al. supra*). For example, a Northern transfer may be used for the detection of the desired mRNA directly. In brief, the mRNA is isolated from a given cell sample using, for example, an acid guanidinium-phenol-chloroform extraction method. The mRNA is then electrophoresed to separate the mRNA species and the mRNA is transferred from the gel to a nitrocellulose membrane. As with the Southern blots, labeled probes are used to identify and/or quantify the target mRNA.

In another preferred embodiment, the gene transcript can be measured using amplification (e.g. PCR) based methods as described above for directly assessing copy number of the target sequences.

Expression of *C. pneumoniae* proteins

The nucleic acids disclosed here can be used for recombinant expression of the proteins. In these methods, the nucleic acids encoding the proteins of interest are introduced into suitable host cells, followed by induction of the cells to produce large amounts of the protein. The invention relies on routine techniques in the field of recombinant genetics, well known to those of ordinary skill in the art. A basic text disclosing the general methods of use in this invention is Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989).

Standard transfection methods are used to produce prokaryotic, mammalian, yeast or insect cell lines which express large quantities of the desired

polypeptide, which is then purified using standard techniques (*see, e.g., Colley et al., J. Biol. Chem.* 264:17619-17622, 1989; *Guide to Protein Purification, supra*).

The nucleotide sequences used to transfect the host cells can be modified to yield *Chlamydia* polypeptides with a variety of desired properties. For example, the polypeptides can vary from the naturally-occurring sequence at the primary structure level by amino acid, insertions, substitutions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

The amino acid sequence variants can be prepared with various objectives in mind, including facilitating purification and preparation of the recombinant polypeptide. The modified polypeptides are also useful for modifying plasma half life, improving therapeutic efficacy, and lessening the severity or occurrence of side effects during therapeutic use. The amino acid sequence variants are usually predetermined variants not found in nature but exhibit the same immunogenic activity as naturally occurring protein. In general, modifications of the sequences encoding the polypeptides may be readily accomplished by a variety of well-known techniques, such as site-directed mutagenesis (*see Gillman & Smith, Gene* 8:81-97 (1979); Roberts *et al., Nature* 328:731-734 (1987)). One of ordinary skill will appreciate that the effect of many mutations is difficult to predict. Thus, most modifications are evaluated by routine screening in a suitable assay for the desired characteristic. For instance, the effect of various modifications on the ability of the polypeptide to elicit a protective immune response can be easily determined using *in vitro* assays. For instance, the polypeptides can be tested for their ability to induce lymphoproliferation, T cell cytotoxicity, or cytokine production using standard techniques.

The particular procedure used to introduce the genetic material into the host cell for expression of the polypeptide is not particularly critical. Any of the well known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, spheroplasts, electroporation, liposomes, microinjection, plasmid vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see Sambrook et al., supra*). It is only necessary that the particular procedure utilized be capable of successfully introducing at least one gene into the host cell which is capable of expressing the gene.

Any of a number of well known cells and cell lines can be used to express the polypeptides of the invention. For instance, prokaryotic cells such as *E. coli* can be used. Eukaryotic cells include, yeast, Chinese hamster ovary (CHO) cells, COS cells, and insect cells.

5 The particular vector used to transport the genetic information into the cell is also not particularly critical. Any of the conventional vectors used for expression of recombinant proteins in prokaryotic and eukaryotic cells may be used. Expression vectors for mammalian cells typically contain regulatory elements from eukaryotic viruses.

10 The expression vector typically contains a transcription unit or expression cassette that contains all the elements required for the expression of the polypeptide DNA in the host cells. A typical expression cassette contains a promoter operably linked to the DNA sequence encoding a polypeptide and signals required for efficient polyadenylation of the transcript. The term "operably linked" as used herein refers to linkage of a
15 promoter upstream from a DNA sequence such that the promoter mediates transcription of the DNA sequence. The promoter is preferably positioned about the same distance from the heterologous transcription start site as it is from the transcription start site in its natural setting. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

20 Following the growth of the recombinant cells and expression of the polypeptide, the culture medium is harvested for purification of the secreted protein. The media are typically clarified by centrifugation or filtration to remove cells and cell debris and the proteins are concentrated by adsorption to any suitable resin or by use of ammonium sulfate fractionation, polyethylene glycol precipitation, or by ultrafiltration.

25 Other routine means known in the art may be equally suitable. Further purification of the polypeptide can be accomplished by standard techniques, for example, affinity chromatography, ion exchange chromatography, sizing chromatography, His₆ tagging and Ni-agarose chromatography (as described in Dobeli *et al.*, *Mol. and Biochem. Parasit.* 41:259-268 (1990)), or other protein purification techniques to obtain homogeneity. The
30 purified proteins are then used to produce pharmaceutical compositions, as described below.

An alternative method of preparing recombinant polypeptides useful as vaccines involves the use of recombinant viruses (e.g., vaccinia). Vaccinia virus is grown

in suitable cultured mammalian cells such as the HeLa S3 spinner cells, as described by Mackett *et al.*, in *DNA cloning Vol. II: A practical approach*, pp. 191-211 (Glover, ed.).

Antibody Production

The proteins of the present invention can be used to produce antibodies specifically reactive with *C pneumoniae* antigens. If isolated proteins are used, they may be recombinantly produced or isolated from *Chlamydia* cultures. Synthetic peptides made using the protein sequences may also be used.

Methods of production of polyclonal antibodies are known to those of skill in the art. In brief, an immunogen, preferably a purified protein, is mixed with an adjuvant and animals are immunized. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera is prepared. Further fractionation of the antisera to enrich for antibodies reactive to *Chlamydia* proteins can be done if desired (*see* Harlow & Lane, *Antibodies: A Laboratory Manual* (1988)).

Polyclonal antisera are used to identify and characterize *Chlamydia* in the tissues of patients using, for instance, *in situ* techniques and immunoperoxidase test procedures described in Anderson *et al. JAVMA* 198:241 (1991) and Barr *et al. Vet. Pathol.* 28:110-116 (1991).

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (*see* Kohler & Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host.

Monoclonal antibodies produced in such a manner are used, for instance, in ELISA diagnostic tests, immunoperoxidase tests, immunohistochemical tests, for the *in vitro* evaluation of spirochete invasion, to select candidate antigens for vaccine development, protein isolation, and for screening genomic and cDNA libraries to select appropriate gene sequences.

Immunodiagnostic detection of *C. pneumoniae* infections

The present invention also provides methods for detecting the presence or absence of *C. pneumoniae*, or antibodies reactive with it, in a biological sample. For instance, antibodies specifically reactive with *Chlamydia* can be detected using either
5 *Chlamydia* proteins or the isolates described here. The proteins and isolates can also be used to raise specific antibodies (either monoclonal or polyclonal) to detect the antigen in a sample. In addition, the nucleic acids disclosed and claimed here can be used to detect *Chlamydia*-specific sequences using standard hybridization techniques.

For a review of immunological and immunoassay procedures in general,
10 see *Basic and Clinical Immunology* (Stites & Terr ed., 7th ed. 1991)). The immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in *Enzyme Immunoassay* (Maggio, ed., 1980); Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology* (1985)). For instance, the proteins and antibodies disclosed here are conveniently used in ELISA, immunoblot analysis and
15 agglutination assays.

In brief, immunoassays to measure anti-*Chlamydia* antibodies or antigens can be either competitive or noncompetitive binding assays. In competitive binding assays, the sample analyte (e.g., anti-*Chlamydia* antibodies) competes with a labeled analyte (e.g., anti-*Chlamydia* monoclonal antibody) for specific binding sites on a capture
20 agent (e.g., isolated *Chlamydia* protein) bound to a solid surface. The concentration of labeled analyte bound to the capture agent is inversely proportional to the amount of free analyte present in the sample.

Noncompetitive assays are typically sandwich assays, in which the sample analyte is bound between two analyte-specific binding reagents. One of the binding
25 agents is used as a capture agent and is bound to a solid surface. The second binding agent is labelled and is used to measure or detect the resultant complex by visual or instrument means.

A number of combinations of capture agent and labelled binding agent can be used. For instance, an isolated *Chlamydia* protein or culture can be used as the
30 capture agent and labelled anti-human antibodies specific for the constant region of human antibodies can be used as the labelled binding agent. Goat, sheep and other non-human antibodies specific for human immunoglobulin constant regions (e.g., γ or μ) are

04430103 111100

well known in the art. Alternatively, the anti-human antibodies can be the capture agent and the antigen can be labelled.

Various components of the assay, including the antigen, anti-*Chlamydia* antibody, or anti-human antibody, may be bound to a solid surface. Many methods for immobilizing biomolecules to a variety of solid surfaces are known in the art. For instance, the solid surface may be a membrane (e.g., nitrocellulose), a microtiter dish (e.g., PVC or polystyrene) or a bead. The desired component may be covalently bound or noncovalently attached through nonspecific bonding.

Alternatively, the immunoassay may be carried out in liquid phase and a variety of separation methods may be employed to separate the bound labeled component from the unbound labelled components. These methods are known to those of skill in the art and include immunoprecipitation, column chromatography, adsorption, addition of magnetizable particles coated with a binding agent and other similar procedures.

An immunoassay may also be carried out in liquid phase without a separation procedure. Various homogeneous immunoassay methods are now being applied to immunoassays for protein analytes. In these methods, the binding of the binding agent to the analyte causes a change in the signal emitted by the label, so that binding may be measured without separating the bound from the unbound labelled component.

Western blot (immunoblot) analysis can also be used to detect the presence of antibodies to *Chlamydia* in the sample. This technique is a reliable method for confirming the presence of antibodies against a particular protein in the sample. The technique generally comprises separating proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the separated proteins. This causes specific target antibodies present in the sample to bind their respective proteins. Target antibodies are then detected using labeled anti-human antibodies.

The immunoassay formats described above employ labelled assay components. The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. A wide variety of labels may be used. The component may be labelled by any one of several methods. Traditionally a radioactive label incorporating ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P was used. Non-radioactive labels

include ligands which bind to labelled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies which can serve as specific binding pair members for a labelled ligand. The choice of label depends on sensitivity required, ease of conjugation with the compound, stability requirements, and available instrumentation.

5 Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidoreductases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, etc. Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, e.g., luminol. For a review of various labelling or
10 signal producing systems which may be used, see U.S. Patent No. 4,391,904, which is incorporated herein by reference.

 Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (e.g., biotin) is covalently bound to the molecule. The ligand then binds to an anti-ligand (e.g., streptavidin) molecule which is either inherently detectable or
15 covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. A number of ligands and anti-ligands can be used. Where a ligand has a natural anti-ligand, for example, biotin, thyroxine, and cortisol, it can be used in conjunction with the labelled, naturally occurring anti-ligands. Alternatively, any haptenic or antigenic compound can be used in combination with an
20 antibody.

 Some assay formats do not require the use of labelled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need be labelled and the presence of
25 the target antibody is detected by simple visual inspection.

Pharmaceutical Compositions

 The peptides or antibodies (typically monoclonal antibodies) of the present invention and pharmaceutical compositions thereof are useful for administration to mammals, particularly humans, to treat and/or prevent *Chlamydia* infections. Suitable
30 formulations are found in *Remington's Pharmaceutical Sciences*, Mack Publishing Company, Philadelphia, PA, 17th ed. (1985).

The immunogenic peptides or antibodies of the invention are administered prophylactically or to an individual already suffering from the disease. The peptide compositions are administered to a patient in an amount sufficient to elicit an effective immune response to *Chlamydia*. An effective immune response is one that inhibits infection. An amount adequate to accomplish this is defined as "therapeutically effective dose" or "immunogenically effective dose." Amounts effective for this use will depend on, e.g., the peptide composition, the manner of administration, the stage and severity of the disease being treated, the weight and general state of health of the patient, and the judgment of the prescribing physician, but generally range for the initial immunization (that is for therapeutic or prophylactic administration) from about 0.1 mg to about 1.0 mg per 70 kilogram patient, more commonly from about 0.5 mg to about 0.75 mg per 70 kg of body weight. Boosting dosages are typically from about 0.1 mg to about 0.5 mg of peptide using a boosting regimen over weeks to months depending upon the patient's response and condition. A suitable protocol would include injection at time 0, 4, 2, 6, 10 and 14 weeks, followed by further booster injections at 24 and 28 weeks.

For therapeutic use, administration should begin at the first sign of infection. This is followed by boosting doses until at least symptoms are substantially abated and for a period thereafter. In some circumstances, loading doses followed by boosting doses may be required. The resulting immune response helps to cure or at least partially arrest symptoms and/or complications. Vaccine compositions containing the peptides are administered prophylactically to a patient susceptible to or otherwise at risk of the infection.

The pharmaceutical compositions (containing either peptides or antibodies) are intended for parenteral or oral administration. Preferably, the pharmaceutical compositions are administered parenterally, e.g., subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration which comprise a solution of the immunogenic polypeptides dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g., water, buffered water, 0.4% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile solution prior to administration. The compositions may contain

pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

5 The compositions may also comprise carriers to enhance the immune response. Useful carriers are well known in the art, and include, e.g., KLH, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, polyamino acids such as poly(lysine:glutamic acid), influenza, hepatitis B virus core protein, hepatitis B virus recombinant vaccine and the like.

10 For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as
15 those carriers previously listed, and generally 10-95% of active ingredient, that is, one or more peptides of the invention, and more preferably at a concentration of 25%-75%.

 As noted above, the peptide compositions are intended to induce an immune response to *Chlamydia*. Thus, compositions and methods of administration suitable for maximizing the immune response are preferred. For instance, peptides may
20 be introduced into a host, including humans, linked to a carrier or as a homopolymer or heteropolymer of active peptide units from various *Chlamydia* proteins disclosed here. Alternatively, a "cocktail" of polypeptides can be used. A mixture of more than one polypeptide has the advantage of increased immunological reaction and, where different peptides are used to make up the polymer, the additional ability to induce antibodies to a
25 number of epitopes.

 The compositions also include an adjuvant. As used here, number of adjuvants are well known to one skilled in the art. Suitable adjuvants include incomplete Freund's adjuvant, alum, aluminum phosphate, aluminum hydroxide,
N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP),
30 N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl

lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against the immunogenic peptide.

5 The concentration of immunogenic peptides of the invention in the pharmaceutical formulations can vary widely, i.e. from less than about 0.1%, usually at or at least about 2% to as much as 20% to 50% or more by weight, and will be selected primarily by fluid volumes, viscosities, etc., in accordance with the particular mode of administration selected.

10 The peptides of the invention can also be expressed by attenuated viral hosts, such as vaccinia or fowlpox. This approach involves the use of vaccinia virus as a vector to express nucleotide sequences that encode the peptides of the invention. Upon introduction into a host, the recombinant vaccinia virus expresses the immunogenic peptide, and thereby elicits an immune response. Vaccinia vectors and methods useful in immunization protocols are described in, e.g., U.S. Patent No. 4,722,848. Another vector
15 is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al. (*Nature* 351:456-460 (1991)). A wide variety of other vectors useful for therapeutic administration or immunization of the peptides of the invention, e.g., *Salmonella typhi* vectors and the like, will be apparent to those skilled in the art from the description herein.

20 The DNA encoding one or more of the peptides of the invention can also be administered to the patient. This approach is described, for instance, in Wolff et al., *Science* 247: 1465-1468 (1990) as well as U.S. Patent Nos. 5,580,859 and 5,589,466.

In order to enhance serum half-life, the peptides may also be encapsulated, introduced into the lumen of liposomes, prepared as a colloid, or other conventional
25 techniques may be employed which provide an extended serum half-life of the peptides. A variety of methods are available for preparing liposomes, as described in, e.g., Szoka et al., *Ann. Rev. Biophys. Bioeng.* 9:467 (1980), U.S. Pat. Nos. 4,235,871, 4,501,728 and 4,837,028.

30

EXAMPLES

The following examples are offered to illustrate, but no to limit the claimed invention.

Example 1:

This example describes comparison of the *C. pneumoniae* genome disclosed here and the, previously sequenced, *C. trachomatis* genome (Stephens, *et al. Science* 282:754-759 (1998)).

The apparent low level of DNA homology between *C. trachomatis* and *C. pneumoniae* (Campbell, *et al., J. Clin. Microbiol.* 25:1911-1916 (1987)) yet analogous cell structures and developmental cycles, predicts that comparative analysis of the two genomes will significantly enhance the understanding of both pathogens. Identification of genes that are present in one species but not the other are of particular importance for the mutually exclusive biological, virulence and pathogenesis capabilities of each.

10 Identification of genes shared between the two species strongly supports the requirement for these capabilities in a biological system that has, over its long-term association with mammalian host cells, evolved to reduce the metabolic capacities while optimizing survival, growth and transmission of these unique pathogens.

The previously sequenced *C. trachomatis* genome contains 1,042,519 nucleotides and 875 likely protein-coding genes. Similarity searching permitted the inferred functional assignment of sequences 636 (60%) genes disclosed here and 251 (23%) are similar to hypothetical genes for other bacterial organisms including those for *C. trachomatis*. The remaining 186 (17%) genes are not homologous to sequences deposited in GenBank.. Seventy *C. trachomatis* genes are not represented in the *C. pneumoniae* genome. These are contained within blocks consisting of 2-17 genes and 19 single genes. Of the 70 *C. trachomatis* genes without homologs in *C. pneumoniae*, 60 are classified as encoding hypothetical proteins. The remaining genes not represented in *C. pneumoniae* consist of the tryptophan operon (*trpA,B,R*), *trpC*, two predicted thiol protease genes, and 4 genes assigned to the phospholipase-D superfamily.

25 It is evident that there is a high level of functional conservation between *C. pneumoniae* and *C. trachomatis* as orthologs to *C. trachomatis* genes were identified for 859 (80%) of the predicted coding sequences for *C. pneumoniae*. The level of similarity for individual encoded proteins spans a wide spectrum (22-95% amino acid identity) with an average of 62% amino acid identity between orthologs from the two species. The

30 percent amino acid identity between orthologous chlamydial proteins is similar among functional groups with the highest for proteins associated with translation and the lowest for proteins whose function in chlamydiae is uncharacterized and not related to proteins encoded by other organisms. The gene order of the homologous set of genes in *C.*

pneumoniae shows reorganization relative to the genome of *C. trachomatis*; however, there is a high level of synteny for the gene organization of the two genomes. We identified thirty-nine blocks of 2 or more genes whose gene organization is colinear with homologs to *C. trachomatis*, although some of these are inverted. The distribution of genome reorganization is not evenly distributed on the chromosome as the region between *C. pneumoniae* coding sequences 0130-0300 contains substantially more reorganization than other areas of the genome. This region coincides with the predicted chromosome replication terminus.

We identified orthologs of enzymes characterized in other bacteria that account for the essential requirements for DNA replication, repair, transcription and translation including two predicted DNA helicases of the Swi2/Snf2 family found in *C. trachomatis*. Similar to *C. trachomatis*, alternative sigma subunits for RNA polymerase, σ^{28} and σ^{54} , were identified in addition to anti- σ regulatory system factors RsbV, a RsbW-like single-domain histidine kinase, and a RsbU-like protein phosphatase. These findings suggest that the fundamental mechanisms of transcriptional regulation are conserved among *Chlamydia*. The *C. trachomatis* proteins containing SET and SWIB domains, and a SWIB domain fused to the C-terminus of the chlamydial topoisomerase I, not identified outside eukaryotes, are found in *C. pneumoniae* supporting their possible role in the chromatin condensation-decondensation characteristic of the biologically unique chlamydial developmental cycle.

The central metabolic pathways inferred from the *C. pneumoniae* genome sequence are the same as those identified for *C. trachomatis*. *C. pneumoniae* has a glycolytic pathway and a linked tricarboxylic acid cycle, although likely functional, is incomplete as genes for citrate synthase, aconitase, and isocitrate dehydrogenase were not identified. *C. pneumoniae* has a complete glycogen synthesis and degradation system supporting a role for glycogen synthesis and utilization of glucose-derivatives in chlamydial metabolism. Genes encoding essential functions in aerobic respiration are present and electron flux may be supported by pyruvate, succinate, glycerol-3-phosphate, and NADH dehydrogenases, NADH-ubiquinone oxidoreductase and cytochrome oxidase. *C. pneumoniae* also contains the V (vacuolar)-type ATPase operon and the two ATP translocases found in *C. trachomatis*.

The type-III secretion virulence system required for invasion by several pathogenic bacteria and found in the *C. trachomatis* genome in three chromosomal

locations is also present in the *C. pneumoniae* genome. Each of the components is conserved and their relative genomic contexts are conserved. Genes such as a predicted serine/threonine protein kinase and other genes physically linked to genes encoding structural components of the type-III secretion apparatus, but without identified
5 homologs, are also highly similar between the two species suggesting the functional roles in modifying cellular biology are fundamentally conserved.

Chlamydia-encoded proteins that are not found in chlamydial organisms but localized to the intracellular chlamydial inclusion membrane are likely essential for the unique intracellular biology and perhaps differences in inclusion morphology
10 observed between species of *Chlamydia*. Several such proteins, termed IncA,B&C, have been characterized for a *C. psittaci* strain (Rockey, *et al. Mol. Microbiol.* 15:617-626 (1995); Rockey *et al. Infect. Immun.* 62:106-112 (1994)). *C. pneumoniae* and *C. trachomatis* encode orthologs to *C. psittaci* IncB and IncC and *C. trachomatis* also contains an ortholog to IncA. *C. pneumoniae* contains two genes that encode proteins
15 with similarity to IncA (CPn0186 and CPn0585), although the level of homology is low suggesting analogous but possibly altered functions.

The tryptophan biosynthesis operon (*trpA*, *trpB*, *trpR*) and *trpC* identified in *C. trachomatis* is conspicuously missing in the *C. pneumoniae* genome. This represents the entire repertoire of genes associated with tryptophan biosynthesis identified
20 in *C. trachomatis*. Seventeen genes adjacent to the *C. trachomatis* tryptophan operon also were not found in the *C. pneumoniae* genome. This region is the single largest loss of a contiguous genomic segment and includes 4 HKD superfamily encoding genes that encompass a family of proteins related to endonuclease and phospholipase D. These findings may be important for the ability of *Chlamydia* to persist in their hosts and cause
25 disease by eliciting potent, focal and persistent inflammatory responses thought to be essential for pathogenesis.

The *C. pneumoniae* genome contains 187,711 additional nucleotides compared to the *C. trachomatis* genome, and the 214 coding sequences not found in *C. trachomatis* account for most of the increased genome size. Eighty-eight of these genes
30 are found in blocks of >10 genes (11-30 genes/block), 41 are single genes, and the remainder are partnered with at least one other gene. Based upon the observation that ~70% of all the *C. pneumoniae* genes have an identifiable homolog in GenBank, exclusive of *C. trachomatis*, it would be expected that over 150 of the 214 genes should

have a homolog in GenBank, many associated with a function. However, only 28 coding sequences have similarity to genes from other organisms. Thus the majority of the genes that are mutually exclusive of *C. trachomatis* (186 of 214), and the 60 of 70 *C. trachomatis* genes that lacked an identifiable homolog in *C. pneumoniae*, do not have detectable homologs to genes from other organisms. We predict that most of the unique genes are essential for specific attributes that define the differential biology, tropism and pathogenesis of *C. trachomatis* and *C. pneumoniae*. Moreover, this suggests that *C. pneumoniae* has more unique biological (i.e., virulence) capacity than *C. trachomatis*. The ability of *C. pneumoniae* to be more invasive and survive in a broader range of host cell types than *C. trachomatis* is consistent with this hypothesis. Not all of the differences in biological capacity may be associated with mutually exclusive genes. One explanation for the significantly lower level of homology between protein sequences assigned as having *C. pneumoniae* and *C. trachomatis* orthologs but no identifiable orthologs in other organisms is that this set of proteins is not only associated with biological requirements specific for *Chlamydia* but this polymorphism may account for differential biology between the two species. The determination of the genome sequence from a representative of the *C. psittaci* group will precisely delineate those genes that are mutually exclusive and specific for each species.

The major functionally identifiable addition to the *C. pneumoniae* genome is a large expansion of genes encoding a new family of chlamydial polymorphic membrane proteins (Pmp), alone representing 22% of the increased coding capacity. While the *C. trachomatis* genome has 9 *pmp* genes, remarkably the *C. pneumoniae* genome contains 21 *pmp* genes. Most of these genes appear to be amplified in two regions of the genome with three stand-alone genes. Interestingly one of the stand-alone genes is most closely related to the *C. trachomatis pmpD* which is the only stand-alone *pmp* gene in the *C. trachomatis* genome and it is located with the same relative genomic context, suggesting an essential and conserved function for this paralog. Six Pmp-coding genes are presumably not functional as five contain predicted coding frame-shifts and one is truncated. The amplification of this gene family and the confidently predicted frame-shifts suggest a specific molecular mechanism to promote functional or antigenic diversity. The biological role of this protein family remains enigmatic, although at least one of the proteins in *C. psittaci* related to this family is exposed on the chlamydial surface.

While a function could not be assigned for most of the unique *C. pneumoniae* genes, several have significant similarity to genes from other organisms. Functional assignments could be made for genes encoding GMP synthetase, IMP dehydrogenase, UMP synthase, uridine kinase, biotin synthase pathway proteins, methylthioadenosine nucleosidase, a DNA glycosylase and aromatic amino acid hydroxylase. Thus a complete pathway was identified for biotin biosynthesis. The additional purine and pyrimidine salvage pathway genes presumably reflect metabolic limitations in one of the cell types that *C. pneumoniae* infects or differences in the ability of *C. pneumoniae* to transport precursor nucleosides or nucleotides.

The addition of aromatic amino acid hydroxylase in *C. pneumoniae* is intriguing especially in light of the loss of tryptophan biosynthetic genes and the inability to synthesize other amino acids including phenylalanine. Aromatic amino acid hydroxylases include three distinct enzymes that function to receptively oxidize phenylalanine to tyrosine, tyrosine to Dopa, and tryptophan to 5-hydroxytryptophan and serotonin. Although the chlamydial protein is similar to proteins of this family and incrementally more closely related to tryptophan hydroxylase, its specific function could not be confidently predicted. We hypothesize that it may be involved in *C. pneumoniae* virulence. Tryptophan hydroxylase has not been previously identified in bacteria and the origin of the chlamydial gene appears to be from eukaryotes. The functional role of an aromatic amino acid hydroxylase for *C. pneumoniae* is linked to the unique intracellular biology of this organism and may represent a key contribution to *C. pneumoniae* persistence and pathogenesis.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

Table 1 provides functional assignments of *C. pneumoniae* nonprotein-encoding genomic sequences. Table 2 provides functional assignments of protein coding sequences. Table 3 provides the amino acid sequences of the proteins corresponding to the coding sequences.

[illegible]36

TABLE 2

Gene #	From	To	Strand	Gene Function (C. trachomatis ortholog in parentheses)
CPn0001	282	4	R	CT001 hypothetical protein
CPn0002	573	875	F	gatC-Glu-tRNA Gln Amidotransferase (C subunit)-(CT002)
CPn0003	895	2370	F	gatA-Glu tRNA Gln Amidotransferase-(CT003)
CPn0004	2370	3122	F	gatB-(Pct112) Gln tRNA Gln Amidotransferase (B Subunit)-(CT004)
CPn0005	4127	6892	F	pmp_1-Polymorphic Outer Membrane Protein G Family
CPn0006	7293	7141	R	
CPn0007	7605	10496	F	
CPn0008	10975	11685	F	
CPn0009	11815	13119	F	
CPn0010	13435	14325	F	
CPn0010	14379	15746	F	frame-shift with 0010
CPn0011	15892	16614	F	
CPn0012	16644	18212	F	
CPn0013	18584	21106	F	pmp_2-Polymorphic Outer Membrane Protein G Family
CPn0014	21392	21922	F	pmp_3-Polymorphic Outer Membrane Protein G Family
CPn0015	21835	24174	F	pmp_3-PMP_3 (frame-shift with 0014)
CPn0016	24416	26188	F	pmp_4-Polymorphic Outer Membrane Protein G Family
CPn0017	26094	27170	F	pmp_4-PMP_4 (frame-shift with 0016)
CPn0018	27522	29003	F	pmp_5-Polymorphic Outer Membrane Protein G Family
CPn0019	29007	30356	F	pmp_5-PMP_5 (frame-shift with 0018)
CPn0020	32687	30603	F	Predicted OMP [leader (14) peptide: outer membrane]-(CT351)
CPn0021	34410	32707	R	Predicted OMP [leader (19) peptide]-(CT350)
CPn0022	34982	34395	R	maf-(CT349)
CPn0023	36603	35014	R	yjjK/alr-ABC Transporter Protein ATPase-(CT348)
CPn0024	37596	36661	F	xerC-Integrase/recombinase-(CT347)
CPn0025	38604	37684	R	elaC/atsA-Sulphohydrolase/Glycosulfatase-(CT346)
CPn0026	39625	38762	R	CT345 hypothetical protein-(CT345)
CPn0027	42234	39778	R	lon-Lon ATP-dependent Protease-(CT344)
CPn0028	43325	42543	R	
CPn0029	43755	43390	R	
CPn0030	43891	44529	F	gcp_1-O-Sialoglycoprotein Endopeptidase_1-(CT343)
CPn0031	44711	44884	F	rs21-S21 Ribosomal Protein-(CT342)
CPn0032	44923	46098	F	dnaJ-Heat Shock Protein J-(CT341)
CPn0033	46138	48171	F	pdhA&B/odhA&odhB-(pyruvate) Oxoisovalerate Dehydrogenase Alpha & Beta Fusion-(CT340)
CPn0034	49457	48210	R	
CPn0035	51029	49569	R	CT339 hypothetical protein
CPn0036	51002	51796	F	CT338 hypothetical protein
CPn0037	51792	52115	F	ptsH-PTS Phosphocarrier Protein Hpr-(CT337)
CPn0038	52119	53831	F	ptsI-PTS PEP Phosphotransferase-(CT336)
CPn0039	54250	53963	R	ybaB-(CT335)
CPn0040	55643	54318	R	dnaX_1-DNA Pol III Gamma and Tau_1-(CT334)
CPn0041	55996	57342	F	
CPn0042	57403	58182	F	
CPn0043	58447	60372	F	
CPn0044	60419	60778	F	
CPn0045	61069	62790	F	
CPn0046	62790	63263	F	
CPn0047	63455	63652	F	
CPn0048	63687	65801	F	*yqfF-Bs conserved hypothetical IM protein
CPn0049	66296	65817	R	
CPn0050	66813	66499	R	
CPn0051	66833	67111	F	
CPn0052	68005	67304	R	hemC-Porphobilinogen Deaminase-(CT299)
CPn0053	69344	67986	R	sms-Sms Protein-(CT298)
CPn0054	70023	69313	R	rnc-Ribonuclease III-(CT297)
CPn0055	70129	70590	F	CT296 hypothetical protein
CPn0056	70953	72746	F	mrsA-Phosphomannomutase-(CT295)
CPn0057	72934	73554	F	sodM-Superoxide Dismutase (Mn)-(CT294)
CPn0058	73639	74562	F	accD-AcCoA Carboxylase/Transferase Beta-(CT293)
CPn0059	74616	75050	F	dut-dUTP Nucleotidohydrolase-(CT292)
CPn0060	75055	75528	F	ptsN_1-PTS IIA Protein-(CT291)
CPn0061	75534	76208	F	ptsN_2-PTS IIA Protein + HTH DNA-Binding Domain-(CT290)
CPn0062	76308	77690	F	CT289 hypothetical protein
CPn0063	78112	78267	F	
CPn0064	78346	78576	F	
CPn0065	78924	80651	F	CT288 hypothetical protein
CPn0066	80925	82655	F	

CPn0067	82953	84053	F	
CPn0068	84903	84331	R	CT360 hypothetical protein
CPn0069	85236	87086	F	
CPn0070	87378	87208	R	
CPn0071	88045	87599	R	CT325 hypothetical protein
CPn0072	89061	88057	R	CT324 hypothetical protein
CPn0073	89356	89574	F	infA-Initiation Factor IF-1-(CT323)
CPn0074	89774	90955	F	tufA-Elongation Factor Tu-(CT322)
CPn0075	91102	91350	F	secE-preprotein translocase-(CT321)
CPn0076	91358	91903	F	nusG-Transcriptional Antitermination-(CT320)
CPn0077	92013	92435	F	rl11-L11 Ribosomal Protein-(CT319)
CPn0078	92465	93160	F	rl1-L1 Ribosomal Protein-(CT318)
CPn0079	93179	93688	F	rl10-L10 Ribosomal Protein-(CT317)
CPn0080	93735	94121	F	rl7-L7/L12 Ribosomal Protein-(CT316)
CPn0081	94261	98016	F	rpoB-RNA Polymerase Beta-(CT315)
CPn0082	98043	102221	F	rpoC-RNA Polymerase Beta'-(CT314)
CPn0083	102332	103312	F	tal-Transaldolase-(CT313)
CPn0084	103362	103751	F	predicted ferredoxin-(CT312)
CPn0085	104506	103766	R	CT311 hypothetical protein
CPn0086	104904	105527	F	atpE-ATP Synthase Subunit E-(CT310)
CPn0087	105579	106376	F	CT309 hypothetical protein
CPn0088	106373	108145	F	atpA-ATP Synthase Subunit A-(CT308)
CPn0089	108153	109466	F	atpB-ATP Synthase Subunit B-(CT307)
CPn0090	109454	110080	F	atpD-ATP Synthase Subunit D-(CT306)
CPn0091	110074	112053	F	atpI-ATP Synthase Subunit I-(CT305)
CPn0092	112151	112573	F	atpK-ATP Synthase Subunit K-(CT304)
CPn0093	112509	113015	F	CT303 hypothetical protein
CPn0094	113152	115971	F	vals-Valyl tRNA Synthetase-(CT302)
CPn0095	116037	118790	F	pknD-S/T Protein Kinase-(CT301)
CPn0096	124314	118837	R	uvrA-Excinuclease ABC Subunit A-(CT333)
CPn0097	124555	126006	F	pyk-Pyruvate Kinase-(CT332)
CPn0098	127491	126091	R	htrB-Acyltransferase-(CT010)
CPn0099	127593	127865	F	
CPn0100	129141	127882	R	CT011 hypothetical protein
CPn0101	129932	129141	R	ybbP family hypothetical protein-(CT012)
CPn0102	130123	131466	F	cydA-Cytochrome Oxidase Subunit I-(CT013)
CPn0103	131480	132511	F	cydB-Cytochrome Oxidase Subunit II-(CT014)
CPn0104	133875	132676	R	CT017 hypothetical protein
CPn0105	134847	134029	R	CT016 hypothetical protein
CPn0106	135091	136374	F	phoH-ATPase-(CT015)
CPn0107	137162	136392	R	CT058 hypothetical protein_1
CPn0108	137857	137303	R	CT018
CPn0109	138655	141783	F	ileS-Isoleucyl-tRNA Synthetase-(CT019)
CPn0110	143734	141827	R	lepB-Signal Peptidase I-(CT020)
CPn0111	144686	143934	R	CT021 hypothetical protein
CPn0112	144767	145093	F	rl31-L31 Ribosomal Protein-(CT022)
CPn0113	145335	146405	F	pfrA-Peptide Chain Releasing Factor (RF-1)-(CT023)
CPn0114	146398	147261	F	hemK-A/G specific methylase-(CT024)
CPn0115	147279	148622	F	ffh-Signal Recognition Particle GTPase-(CT025)
CPn0116	148616	148972	F	rs16-S16 Ribosomal Protein-(CT026)
CPn0117	148989	150071	F	trmD-tRNA (guanine N-1)-Methyltransferase-(CT027)
CPn0118	150102	150464	F	rl19-L19 Ribosomal Protein-(CT028)
CPn0119	150523	151164	F	rnhB_1-Ribonuclease HII_1-(CT029)
CPn0120	151164	151778	F	gmk-GMP Kinase-(CT030)
CPn0121	151778	152068	F	CT031 hypothetical protein
CPn0122	152071	153723	F	metG-Methionyl-tRNA Synthetase-(CT032)
CPn0123	155969	153774	R	recD_1-Exodeoxyribonuclease V (Alpha Subunit)_1-(CT033)
CPn0124	156614	158068	F	
CPn0125	158096	158605	F	
CPn0126	158809	161085	F	
CPn0127	162143	161130	R	ytff-Cationic Amino Acid Transporter-(CT034)
CPn0128	162277	163053	F	bpl1-Biotin Protein Ligase-(CT035)
CPn0129	163717	163064	R	similarity to CT036
CPn0130	164245	163751	R	
CPn0131	164549	165580	F	
CPn0132	165587	166561	F	
CPn0133	167334	166564	R	CHLPS hypothetical protein-(CT109)
CPn0134	169098	167467	R	groEL_1-HSP-60_1-(CT110)
CPn0135	169448	169143	R	groES-10KDa Chaperonin-(CT111)
CPn0136	171401	169569	R	pepF-Oligopeptidase-(CT112)
CPn0137	172254	171502	R	ybgI-ACR family-(CT108)
CPn0138	174019	172700	R	hemL-Glutamate-1-semialdehyde-2,1-aminomutase-(CT210)

CPn0139	174656	174093	R	yqgE-(CT210)
CPn0140	175110	174673	R	yqdE-(CT212)
CPn0141	175802	175110	R	rpiA-Ribose-5-P Isomerase A-(CT213)
CPn0142	176091	175816	R	
CPn0143	177335	176214	R	*yxjG_Bs_1 Hypothetical Protein
CPn0144	177963	180560	F	clpB-Clp Protease ATPase-(CT113)
CPn0145	180777	182369	F	CT114 hypothetical protein
CPn0146	182613	183095	F	
CPn0147	183225	183671	F	
CPn0148	183846	185702	F	pkn1-S/T Protein Kinase-(CT145)
CPn0149	185715	187700	F	dnlJ-DNA Ligase-(CT146)
CPn0150	187834	192444	F	CT147 hypothetical protein
CPn0151	194142	192625	R	mhpA-Monooxygenase-(CT148)
CPn0152	195265	194318	R	CT149 hypothetical protein
CPn0153	195433	197892	F	leuS-Leucyl tRNA Synthetase-(CT209)
CPn0154	197892	199202	F	gseA-KDO Transferase-(CT208)
CPn0155	199691	199488	R	
CPn0156	200117	199770	R	
CPn0157	200723	200298	R	
CPn0158	201430	200894	R	
CPn0159	201772	201467	R	
CPn0160	203791	202127	R	pfkA_1-Fructose-6-P Phosphotransferase_1-(CT207)
CPn0161	204622	203798	R	predicted acyltransferase family-(CT206)
CPn0162	205828	204803	R	
CPn0163	206026	206394	F	
CPn0164	206498	206998	F	
CPn0165	206998	207582	F	
CPn0166	207630	207962	F	
CPn0167	208306	207977	R	
CPn0168	208641	208417	R	
CPn0169	209501	208710	R	
CPn0170	211026	210025	R	
CPn0171	212435	211149	R	*guaA-GMP Synthase
CPn0172	213177	212440	R	*guaB/impD-Inosine 5'-monophosphate dehydrogenase (C00H-terminal region only)
CPn0173	213987	213715	R	
CPn0174	214257	214724	F	
CPn0175	214898	215275	F	
CPn0176	215286	216518	F	CT153 hypothetical protein
CPn0177	217459	216608	R	
CPn0178	218052	217789	R	
CPn0179	218403	218056	R	
CPn0180	218851	218355	R	
CPn0181	219175	218777	R	
CPn0182	220695	219334	R	accC-Biotin Carboxylase-(CT124)
CPn0183	221195	220695	R	accB-Biotin Carboxyl Carrier Protein-(CT123)
CPn0184	221775	221221	R	efp_1-Elongation Factor P_1-(CT122)
CPn0185	222451	221765	R	rpe/araD-Ribulose-P Epimerase-(CT121)
CPn0186	222899	224068	F	*similarity to Cps Inca_1-(CT119)
CPn0187	224248	225045	F	predicted methylase-(CT133)
CPn0188	225111	226400	F	CT132 hypothetical protein
CPn0189	226400	229825	F	CT131 homolog-(Possible Transmembrane Protein)
CPn0190	229919	231274	F	
CPn0191	231991	231314	R	glnQ-ABC Amino Acid Transporter ATPase-(CT130)
CPn0192	232634	231984	R	glnP-ABC Amino Acid Transporter Permease-(CT129)
CPn0193	233126	232686	R	*argR-Arginine Repressor
CPn0194	233210	234241	F	gcp_2-O-Sialoglycoprotein Endopeptidase_2-(CT197)
CPn0195	234190	235785	F	oppA_1-Oligopeptide Binding Protein_1
CPn0196	235939	237519	F	oppA_2-Oligopeptide Binding Protein_2-(CT198)
CPn0197	237578	238882	F	oppA_3-Oligopeptide Binding Protein_3
CPn0198	239169	240746	F	oppA_4-Oligopeptide Binding Protein_4
CPn0199	241042	241983	F	oppB_1-Oligopeptide Permease_1-(CT199)
CPn0200	242017	242868	F	oppC_1-Oligopeptide Permease_1-(CT200)
CPn0201	242864	243715	F	oppD-Oligopeptide Transport ATPase-(CT201)
CPn0202	243715	244500	F	oppF-Oligopeptide Transport ATPase-(CT202)
CPn0203	245008	245802	F	
CPn0204	245817	246002	F	
CPn0205	246133	246327	F	
CPn0206	246409	247161	F	CT203 hypothetical protein
CPn0207	247208	248617	F	ybhI/sodiT1-Oxoglutarate/Malate Translocator-(CT204)
CPn0208	248953	250602	F	pfkA_2-Fructose-6-P Phosphotransferase_2-(CT205)
CPn0209	251036	251272	F	

CPn0210	252384	251440	R	
CPn0211	252756	252463	R	
CPn0212	254066	252888	R	
CPn0213	254342	254190	R	
CPn0214	255657	254446	R	
CPn0215	257015	255759	R	
CPn0216	257608	257174	R	
CPn0217	257896	258579	F	ypdP-(CT140)
CPn0218	259058	258582	R	
CPn0219	259357	260472	F	tgt-Queuine tRNA Ribosyl Transferase-(CT193)
CPn0220	260696	261238	F	
CPn0221	261657	262064	F	
CPn0222	262504	262842	F	*weak similarity to Bacteriophage CHP1 (Orf4)
CPn0223	262956	263333	F	
CPn0224	263435	263674	F	
CPn0225	263873	264541	F	
CPn0226	264566	264967	F	
CPn0227	265416	265009	R	dsbB-Disulfide bond Oxidoreductase-(CT176)
CPn0228	266110	265412	R	dsbG-Disulfide Bond Chaperone-(CT177)
CPn0229	266328	267560	F	CT178 hypothetical protein
CPn0230	268253	267576	R	CT179 hypothetical protein
CPn0231	268957	268253	R	tauB-ABC Transport ATPase (Nitrate/Fe)-(CT180)
CPn0232	270122	269232	R	*similarity to 5'-Methylthioadenosine / S-Adenosylhomocysteine Nucleosidase
CPn0233	270424	270248	R	
CPn0234	271240	270548	R	CT181 hypothetical protein
CPn0235	271416	272177	F	kdsB-deoxyoctulononic Acid Synthetase-(CT182)
CPn0236	272156	273766	F	pyrG-CTP Synthetase-(CT183)
CPn0237	273762	274214	F	yggF Family-(CT184)
CPn0238	274303	275838	F	zwf-Glucose-6-P Dehydrogenase-(CT185)
CPn0239	275899	276672	F	devB-Glucose-6-P Dehydrogenase (DevB family)-(CT186)
CPn0240	277861	276698	R	
CPn0241	279354	278203	R	
CPn0242	279918	279487	R	
CPn0243	280555	280133	R	
CPn0244	280918	281556	F	adk-Adenylate Kinase-(CT128)
CPn0245	281645	282499	F	ydH-Polysaccharide Hydrolase-Invasin Repeat Family-(CT127)
CPn0246	282952	282551	R	rs9-S9 Ribosomal Protein-(CT126)
CPn0247	283415	282969	R	rl13-L13 Ribosomal Protein-(CT125)
CPn0248	284327	283650	R	ycfV/ybba-ABC Transporter ATPase-(CT152)
CPn0249	285841	284333	R	CT151 hypothetical protein
CPn0250	286057	285902	R	rl33-L33 Ribosomal Protein-(CT150)
CPn0251	286060	287559	F	*conserved hypothetical protein
CPn0252	288112	287576	R	CT144 hypothetical protein (frame-shift with 0253?)
CPn0253	288456	287950	R	CT144 hypothetical protein_1
CPn0254	289262	288459	R	CT143 hypothetical protein_1
CPn0255	290165	289329	R	CT142 hypothetical protein_1
CPn0256	291264	290398	R	CT144 hypothetical protein_2
CPn0257	292127	291267	R	CT143 hypothetical protein_2
CPn0258	292534	292133	R	CT142 hypothetical protein (frame-shift with 0259?)
CPn0259	292986	292441	R	CT142 hypothetical protein_2
CPn0260	294045	293548	R	secA_1-Protein Translocase Subunit_1-(CT141)
CPn0261	294302	295033	F	ydaO-PP-Loop Superfamily ATPase-(CT217)
CPn0262	295091	295933	F	surE-SurE-like Acid Phosphatase-(CT218)
CPn0263	296249	297136	F	yqfU hypothetical protein-(CT221)
CPn0264	297730	297155	R	ubiD-Phenylacrylate Decarboxylase-(CT220)
CPn0265	298620	297730	R	ubiA-Benzoate Octaphenyltransferase-(CT219)
CPn0266	299184	299876	F	
CPn0267	300122	300910	F	
CPn0268	300935	301318	F	
CPn0269	302450	301476	R	Dipeptidase-(CT138)
CPn0270	303325	302468	R	ylwC-SuA5 Superfamily-related Protein-(CT137)
CPn0271	303634	304362	F	Lysophospholipase esterase-(CT136)
CPn0272	305233	304340	R	dnaX_2-DNA Pol III Gamma and Tau_2-(CT187)
CPn0273	305844	305227	R	tdk-Thymidylate Kinase-(CT188)
CPn0274	308353	305852	R	gyrA_1-DNA Gyrase Subunit A_1-(CT189)
CPn0275	310786	308372	R	gyrB_1-DNA Gyrase Subunit B_1-(CT190)
CPn0276	311137	310793	R	CT191 hypothetical protein
CPn0277	311910	311404	R	
CPn0278	312875	312060	R	*conserved outer membrane lipoprotein protein
CPn0279	313537	312875	R	*Possible ABC Transporter Permease Protein
CPn0280	314572	313550	R	dppF-Dipeptide Transporter ATPase-(CT689)

CPn0281	315057	316103	F	dhnA-Predicted 1,6-Fructose Biphosphate Aldolase (dehydrin family)-(CT215)
CPn0282	316126	317529	F	xasA/gadC-Amino Acid Transporter-(CT216)
CPn0283	318497	317532	R	
CPn0284	319045	318551	R	
CPn0285	320595	319051	R	
CPn0286	322059	320650	R	mgtE-Mg++ Transporter (CBS Domain)-(CT194)
CPn0287	324221	322089	R	
CPn0288	325716	324571	R	CT195 hypothetical protein
CPn0289	325812	326996	F	aaaT-Neutral Amino Acid (Glutamate) Transporter-(CT230)
CPn0290	327042	328523	F	Na-dependent Transporter-(CT231)
CPn0291	328667	329194	F	incB-Inclusion Membrane Protein B-(CT232)
CPn0292	329228	329836	F	incC-Inclusion Membrane Protein C-(CT233)
CPn0293	329949	332723	F	CT234 hypothetical protein
CPn0294	333092	333502	F	cAMP-Dependent Protein Kinase Regulatory Subunit-(CT235)
CPn0295	333863	333627	R	acpP-Acyl Carrier Protein-(CT236)
CPn0296	334765	334022	R	fabG-Oxoacyl (Carrier Protein) Reductase-(CT237)
CPn0297	335697	334774	R	fabD-Malonyl Acyl Carrier Transcyclase-(CT238)
CPn0298	336721	335717	R	fabH-Oxoacyl Carrier Protein Synthase III-(CT239)
CPn0299	336816	337415	F	recR-Recombination Protein-(CT240)
CPn0300	337783	340152	F	yaeT-Omp85 Analog-(CT241)
CPn0301	340250	340762	F	(OmpH-Like Outer Membrane Protein)-(CT242)
CPn0302	340787	341866	F	lpxD-UDP Glucosamine N-Acyltransferase-(CT243)
CPn0303	342958	341921	R	CT244 hypothetical protein
CPn0304	343133	344158	F	pdhA/odpA-Pyruvate Dehydrogenase Alpha-(CT245)
CPn0305	344154	345137	F	pdhB/odpB-Pyruvate Dehydrogenase Beta-(CT246)
CPn0306	345145	346431	F	pdhC-Dihydrolipoamide Acetyltransferase-(CT247)
CPn0307	348986	346515	F	glgP-Glycogen Phosphorylase-(CT248)
CPn0308	349234	349596	F	similarity to CT249
CPn0309	350974	349595	R	dnaA_1-Replication Initiation Protein_1-(CT250)
CPn0310	353433	351049	R	60IM-60kDa Inner Membrane Protein-(CT251)
CPn0311	354438	353575	R	lgt-Prolipoprotein Diacylglycerol Transferase-(CT252)
CPn0312	354524	354976	F	CT101 hypothetical protein
CPn0313	354990	355355	F	acpS-Acyl-carrier Protein Synthase-(CT100)
CPn0314	356285	355353	R	trxB-Thioredoxin Reductase-(CT099)
CPn0315	356977	358716	F	rs1-S1 Ribosomal Protein-(CT098)
CPn0316	358820	360121	F	nusA-N Utilization Protein A-(CT097)
CPn0317	360081	362750	F	infB-Initiation Factor-2-(CT096)
CPn0318	362767	363126	F	rbfA-Ribosome Binding Factor A-(CT095)
CPn0319	363175	363879	F	truB-tRNA Pseudouridine Synthase-(CT094)
CPn0320	363860	364783	F	ribF-FAD Synthase-(CT093)
CPn0321	365858	364767	R	yehF-GTP Binding Protein-(CT092)
CPn0322	366249	367328	F	yscU-YopS Translocation Protein U -(CT091)
CPn0323	367331	369460	F	lcrD- Low Calcium Response D-(CT090)
CPn0324	369492	370688	F	lcrE- Low Calcium Response E-(CT089)
CPn0325	370708	371148	F	syncE-Secretion Chaperone-(CT088)
CPn0326	371148	372725	F	malQ-Glucanotransferase-(CT087)
CPn0327	372945	373211	F	rl28-L28 Ribosomal Protein-(CT086)
CPn0328	373241	374992	F	CT085 hypothetical protein
CPn0329	375088	376146	F	Phospholipase D Superfamily [leader (33) peptide]-(CT084)
CPn0330	376675	376202	R	CT083 hypothetical protein
CPn0331	378437	376701	R	CT082 hypothetical protein
CPn0332	378655	378536	R	CHLTR T2 Protein-(CT081)
CPn0333	379090	378800	R	ltuB-(CT080)
CPn0334	379311	379823	F	CT079 similarity
CPn0335	379817	380674	F	fold-Methylene Tetrahydrofolate Dehydrogenase-(CT078)
CPn0336	380650	381591	F	yojL-(CT077)
CPn0337	382027	381575	R	smpB- Small Protein B-(CT076)
CPn0338	382278	383375	F	dnaN-DNA Pol III (beta chain)-(CT075)
CPn0339	383420	384034	F	recF-ABC superfamily ATPase-(CT074)
CPn0340	383842	384156	F	(frame-shift with 0339)
CPn0341	384160	384495	F	(frame-shift with 0340)
CPn0342	384622	385062	F	predicted OMP [leader (19) peptide]-(CT073)
CPn0343	384999	385595	F	(frame-shift with 0342?)
CPn0344	387420	385558	R	yaeL-Metalloprotease-(CT072)
CPn0345	388572	387436	R	yaeM-(CT071)
CPn0346	389675	388704	R	troD/ytgD-Integral Membrane Protein-(CT070)
CPn0347	391021	389678	R	troC/ytgC-Integral Membrane Protein-(CT069)
CPn0348	391803	391027	R	troB/ytgB-ABC transporter ATPase-(CT068)
CPn0349	392770	391790	R	troA/ytgA-Solute Protein Binding Family-(CT067)
CPn0350	393181	393664	F	CT066 hypothetical protein
CPn0351	393888	395432	F	adt_1-ADP/ATP Translocase_1-(CT065)

CPn0352	395574	396830	F	
CPn0353	396893	397135	F	
CPn0354	397167	398507	F	
CPn0355	399889	398591	R	
CPn0356	400459	400109	R	
CPn0357	401317	400469	R	
CPn0358	401751	401578	R	
CPn0359	402012	403817	F	lepA-GTPase-(CT064)
CPn0360	405358	403922	R	gnd-6-Phosphogluconate Dehydrogenase-(CT063)
CPn0361	406647	405382	R	tyrS-tyrosyl tRNA Synthetase-(CT062)
CPn0362	407825	407055	R	fliA/xpsD-Sigma-28/WhiG Family-(CT061)
CPn0363	409688	407943	R	flhA-Flagellar Secretion Protein-(CT060)
CPn0364	409966	410238	F	fer4-Ferredoxin IV-(CT059)
CPn0365	410528	411544	F	
CPn0366	411976	412440	F	
CPn0367	413102	413836	F	
CPn0368	413790	414107	F	
CPn0369	414351	415562	F	CT058 hypothetical protein_2
CPn0370	415800	416912	F	CT058 hypothetical protein_3
CPn0371	417147	417503	F	
CPn0372	417687	418001	F	
CPn0373	418380	420218	F	gcpE-(CT057)
CPn0374	420218	420961	F	CT056 hypothetical protein
CPn0375	421121	421615	F	
CPn0376	421854	422294	F	
CPn0377	423438	422347	R	sucB_1-Dihydrolipoamide Succinyltransferase_1-(CT055)
CPn0378	426168	423445	R	sucA-Oxoglutarate Dehydrogenase-(CT054)
CPn0379	426322	426765	F	CT053 hypothetical protein
CPn0380	426758	427876	F	hemN_1-Coproporphyrinogen III Oxidase_1-(CT052)
CPn0381	429809	428037	R	CT326 similarity
CPn0382	430749	430036	R	yabC/yrpA-SAM-Dependent Methyltransferase-(CT048)
CPn0383	431693	430749	R	CT047 hypothetical protein
CPn0384	432377	431862	R	hctB-Histone-like Protein 2-(CT046)
CPn0385	434018	432522	R	pepA-Leucyl Aminopeptidase A-(CT045)
CPn0386	434525	434046	R	ssb-SS DNA Binding Protein-(CT044)
CPn0387	435196	434699	R	CT043 hypothetical protein
CPn0388	435329	437320	F	glgX-Glycogen Hydrolase (debranching)-(CT042)
CPn0389	438134	437319	R	CT041 hypothetical protein
CPn0390	439144	438134	R	ruvB-Holliday Junction Helicase-(CT040)
CPn0391	439692	439510	R	
CPn0392	439814	440383	F	dcd-dCTP Deaminase-(CT039)
CPn0393	440379	440723	F	CT038 hypothetical protein
CPn0394	440736	441968	F	tlyC_1-CBS Domain protein (Hemolysin Homolog)_1-(CT256)
CPn0395	441964	443175	F	CT257 hypothetical protein
CPn0396	444353	443241	R	yhfo-Nifs-related protein-(CT258)
CPn0397	445115	444381	R	PP2C phosphatase family-(CT259)
CPn0398	445533	445700	F	
CPn0399	445879	446523	F	CT253 hypothetical protein
CPn0400	446536	447306	F	CT254 hypothetical protein
CPn0401	447884	447495	R	CT255 hypothetical protein
CPn0402	448994	447888	R	mutY-Adenine Glycosylase-(CT107)
CPn0403	449015	449710	F	yceC-predicted pseudouridine synthetase family-(CT106)
CPn0404	450887	449871	R	
CPn0405	451739	450966	R	CT105 hypothetical protein
CPn0406	451969	452865	F	fabI-Enoyl-Acyl-Carrier Protein Reductase-(CT104)
CPn0407	453742	452858	R	HAD superfamily hydrolase/phosphatase-(CT103)
CPn0408	454105	454581	F	CT102 hypothetical protein
CPn0409	454645	455127	F	CT260 hypothetical protein
CPn0410	455123	455833	F	dnaQ_1-DNA Pol III Epsilon Chain_1-(CT261)
CPn0411	455833	456609	F	CT262 hypothetical protein
CPn0412	456590	457246	F	CT263 hypothetical protein
CPn0413	459203	457227	R	msbA-Transport ATP Binding Protein-(CT264)
CPn0414	460143	459172	R	accA-AcCoA Carboxylase/Transferase Alpha-(CT265)
CPn0415	461498	460221	R	CT266 hypothetical protein
CPn0416	461856	461557	R	himD/ihfA-Integration Host Factor Alpha-(CT267)
CPn0417	463035	462244	R	amiA-N-Acetylmuramoyl Alanine Amidase-(CT268)
CPn0418	464401	462953	R	murE-N-Acetylmuramoylalananylgutanyl DAP Ligase-(CT269)
CPn0419	466834	464876	R	pbp3- transglycolase/transpeptidase-(CT270)
CPn0420	467108	466824	R	CT271 hypothetical protein
CPn0421	467998	467108	R	yabC-PBP2B Family methyltransferase-(CT272)
CPn0422	468242	468784	F	CT273 hypothetical protein
CPn0423	468791	469216	F	CT274 hypothetical protein

CPn0424	469612	470961	F	dnaA_2-Replication Initiation Factor_4-(CT275)
CPn0425	470980	471564	F	CT276 hypothetical proteins
CPn0426	472111	471536	R	CT277 similarity
CPn0427	472207	473715	F	nqr2-NADH (Ubiquinone) Dehydrogenase-(CT278)
CPn0428	473722	474681	F	nqr3-NADH (Ubiquinone) Oxidoreductase, Gamma-(CT279)
CPn0429	474681	475319	F	nqr4-NADH (Ubiquinone) Reductase 4-(CT280)
CPn0430	475326	476093	F	nqr5-NADH (Ubiquinone) Reductase 5-(CT281)
CPn0431	476483	476151	R	
CPn0432	476816	476514	R	
CPn0433	477273	476929	R	gcsH-Glycine Cleavage System H Protein-(CT282)
CPn0434	479462	477276	R	CT283 hypothetical protein
CPn0435	480902	479475	R	Phospholipase D superfamily [uncleavable leader peptide]-(CT284)
CPn0436	481618	480902	R	lplA-Lipoate Protein Ligase-Like Protein-(CT285)
CPn0437	481816	484350	F	clpC-ClpC Protease-(CT286)
CPn0438	485416	484334	R	ycbF-PP-loop superfamily ATPase-(CT287)
CPn0439	485553	486077	F	
CPn0440	486105	486740	F	
CPn0441	486891	487838	F	CT007 hypothetical protein
CPn0442	488013	488528	F	CT006 hypothetical protein
CPn0443	488729	489979	F	CT005 hypothetical protein
CPn0444	490287	494507	F	pmp_6-Polymorphic Outer Membrane Protein G/I Family
CPn0445	494772	497579	F	pmp_7-Polymorphic Outer Membrane Protein G Family
CPn0446	497626	500415	F	pmp_8-Polymorphic Outer Membrane Protein G Family
CPn0447	500568	503351	F	pmp_9-Polymorphic Outer Membrane Protein G/I Family
CPn0448	504810	503698	R	*yxjG_Bs_2 Hypothetical Protein
CPn0449	507231	505330	R	pmp_10-PMP_10 (Frame-shift with 0451)
CPn0450	508112	507180	R	pmp_10-Polymorphic Outer Membrane Protein G Family
CPn0451	508275	511058	F	pmp_11-Polymorphic Outer Membrane Protein G Family
CPn0452	511319	512860	F	pmp_12-Polymorphic Outer Membrane Protein A/I Family (truncated)
CPn0453	513234	516152	F	pmp_13 -Polymorphic Outer Membrane Protein G Family
CPn0454	516182	519115	F	pmp_14-Polymorphic Outer Membrane Protein H Family
CPn0455	520348	519458	R	
CPn0456	521532	520327	R	
CPn0457	523865	522120	R	
CPn0458	526320	524236	R	
CPn0459	527005	526619	R	
CPn0460	527840	526992	R	
CPn0461	528638	527844	R	
CPn0462	531052	529037	R	
CPn0463	532357	531191	R	
CPn0464	532842	532366	R	
CPn0465	533212	532871	R	
CPn0466	533724	536537	F	pmp_15-Polymorphic Outer Membrane Protein E Family
CPn0467	536633	539434	F	pmp_16-Polymorphic Outer Membrane Protein E Family
CPn0468	539632	540432	F	pmp_17-Polymorphic Outer Membrane Protein E Family
CPn0469	540399	541460	F	pmp_17-Polymorphic Outer Membrane Protein (Frame-shift with 0469)
CPn0470	541357	542532	F	pmp_17-Polymorphic Outer Membrane Protein (Frame-shift with 0470)
CPn0471	542564	545401	F	pmp_18-Polymorphic Outer Membrane Protein E/F Family
CPn0472	547905	545581	R	
CPn0473	549593	548070	R	
CPn0474	551573	549807	R	CT365 hypothetical protein
CPn0475	553844	551685	R	glgB-Glucan Branching Enzyme-(CT866)
CPn0476	554844	553858	R	CT865 hypothetical protein
CPn0477	556106	554844	R	*yqeV_Bs Hypothetical Protein
CPn0478	557625	556210	R	hflX-GTP Binding Protein-(CT379)
CPn0479	558425	557616	R	phnP-Metal Dependent Hydrolase-(CT380)
CPn0480	559303	558650	R	CT383 hypothetical protein
CPn0481	560946	559339	R	
CPn0482	561737	560961	R	artJ-Arginine Periplasmic Binding Protein-(CT381)
CPn0483	561836	564964	F	
CPn0484	564970	565824	F	aroG-Deoxyheptonate Aldolase-(CT382)
CPn0485	566038	566229	F	CT382.1 hypothetical protein
CPn0486	567784	566405	R	*hypothetical proline permease
CPn0487	569740	568112	R	CT384 hypothetical protein
CPn0488	570096	569767	R	hitA-HIT Family Hydrolase-(CT385)
CPn0489	570965	570096	R	CT386 hypothetical protein
CPn0490	571279	573333	F	CT387 hypothetical protein
CPn0491	574352	573336	R	CT389 hypothetical protein
CPn0492	574652	574804	F	
CPn0493	575004	574855	R	
CPn0494	575364	575146	R	
CPn0495	575603	576793	F	aspC-Aspartate Aminotransferase-(CT390)

CPn0496	576793	577812	F	CT391 hypothetical protein
CPn0497	578089	577820	R	CT388 hypothetical protein
CPn0498	579035	578085	R	
CPn0499	580359	579205	R	
CPn0500	580659	582362	F	proS-Prolyl tRNA Synthetase-(CT393)
CPn0501	582457	583650	F	hrcA-MTH Transcriptional Repressor-(CT394)
CPn0502	583650	584201	F	grpE-HSP-70 Cofactor-(CT395)
CPn0503	584234	586213	F	dnaK-HSP-70-(CT396)
CPn0504	586487	588514	F	vacB-ribonuclease family-(CT397)
CPn0505	588519	589106	F	*3-methyladenine DNA glycosylase
CPn0506	589172	589840	F	CT421 hypothetical protein
CPn0507	589961	590122	F	CT421.1 hypothetical protein
CPn0508	590142	590300	F	CT421.2 hypothetical protein
CPn0509	590335	590808	F	(predicted Metalloenzyme)-(CT422)
CPn0510	590813	591973	F	tlyC_2-CBS Domains (Hemolysin homolog)_2-(CT423)
CPn0511	592141	592488	F	rsbV_1-Sigma Regulatory Factor_1-(CT424)
CPn0512	592553	594412	F	CT425 hypothetical protein
CPn0513	594647	595753	F	Fe-S oxidoreductase_1-(CT426)
CPn0514	595729	596520	F	CT427 hypothetical protein
CPn0515	596492	597181	F	ubiE-Ubiquinone Methyltransferase-(CT428)
CPn0516	598814	597255	R	
CPn0517	599631	598795	R	
CPn0518	600803	599832	R	CT429 hypothetical protein
CPn0519	601674	600904	R	dapF-Diaminopimelate Epimerase-(CT430)
CPn0520	602218	601646	R	clpP-CLP Protease-(CT431)
CPn0521	603797	602241	R	glyA-Serine Hydroxymethyltransferase-(CT432)
CPn0522	603987	604655	F	CT433 hypothetical protein
CPn0523	604723	605052	F	
CPn0524	605103	606179	F	
CPn0525	606522	607283	F	CT398 hypothetical protein
CPn0526	608696	607710	R	yrbH-GutQ/KpsF Family Sugar-P Isomerase-(CT399)
CPn0527	609904	608726	R	sucB_2-Dihydrolipoamide Succinyltransferase_2-(CT400)
CPn0528	611162	609921	R	gltT-Glutamate Symport-(CT401)
CPn0529	612259	611165	R	ycaH-ATPase-(CT402)
CPn0530	613254	612460	R	spoU_1-rRNA Methylase_1-(CT403)
CPn0531	614069	613245	R	SAM dependent methyltransferase-(CT404)
CPn0532	614674	614075	R	ribC/risA-Riboflavin Synthase-(CT405)
CPn0533	614930	615385	F	CT406 hypothetical protein
CPn0534	615413	615784	F	dkSA-DnaK Suppressor-(CT407)
CPn0535	615793	616296	F	lspA-Lipoprotein Signal Peptidase-(CT408)
CPn0536	616345	617691	F	dagA_1-D-Ala/Gly Permease_1-(CT409)
CPn0537	617833	618189	F	CT814.1 hypothetical protein
CPn0538	618212	618511	F	CT814 hypothetical protein
CPn0539	618705	621545	F	pmp_19-polymorphic outer membrane protein A Family -(CT412)
CPn0540	621694	626862	F	pmp_20-polymorphic outer membrane protein B Family-(CT413)
CPn0541	627170	628003	F	Solute binding protein (-yebL-Synechocystis Adhesin Homolog)-(CT415)
CPn0542	628003	628737	F	ABC Transporter ATPase-(CT416)
CPn0543	628725	629603	F	(Metal Transport Protein)-(CT417)
CPn0544	630529	629525	R	yhbZ-GTP binding protein-(CT418)
CPn0545	630884	630633	R	rl27-L27 ribosomal protein-(CT419)
CPn0546	631229	630912	R	rl21-L21 Ribosomal Protein-(CT420)
CPn0547	631661	632188	F	ygbB family-(CT434)
CPn0548	633231	632191	R	cysJ-Sulfite Reductase-(CT435)
CPn0549	633569	633255	R	rs10-S10 Ribosomal Protein-(CT436)
CPn0550	635661	633580	R	fusA-Elongation Factor G-(CT437)
CPn0551	636168	635698	R	rs7-S7 Ribosomal Protein-(CT438)
CPn0552	636587	636219	R	rs12-S12 Ribosomal Protein-(CT439)
CPn0553	637747	636812	R	
CPn0554	637854	638141	F	CT440 hypothetical protein
CPn0555	638298	640241	F	tsp-Tail-Specific Protease-(CT441)
CPn0556	640912	640325	R	crpA-15kDa Cysteine-Rich Protein-(CT442)
CPn0557	642861	641194	R	omcB-60kDa Cysteine-Rich Outer Membrane Complex Protein-(CT443)
CPn0558	643300	643031	R	omcA-9kDa-Cysteine-Rich Outer Membrane Complex Lipoprotein-(CT444)
CPn0559	643742	643927	F	CT441.1 hypothetical protein
CPn0560	645612	644098	R	gltX-Glutamyl-tRNA Synthetase-(CT445)
CPn0561	646404	645871	R	euo-CHLPS Euo Protein-(CT446)
CPn0562	648036	646918	R	*CHLPS 43 kDa protein homolog_1
CPn0563	650056	648293	R	recJ-ssDNA Exonuclease-(CT447)
CPn0564	654350	650145	R	secD&secF-Protein Export Proteins SecD/SecF (fusion)-(CT448)
CPn0565	655630	654533	R	CT449 hypothetical protein
CPn0566	656141	656890	F	yaeS family-(CT450)
CPn0567	656894	657817	F	cdaA-Phosphatidate Cytidyltransferase-(CT451)

CPn0568	657817	658464	F	cdsA-Phosphatidate Cytidylyltransferase-(CT452)
CPn0569	658464	659099	F	plsC-Glycerol-3-P Acyltransferase-(CT453)
CPn0570	659107	660789	F	argS-Arginyl tRNA Transferase-(CT454)
CPn0571	662122	660749	R	murA-UDP-N-Acetylglucosamine Transferase-(CT455)
CPn0572	662352	664616	F	CT456 hypothetical protein
CPn0573	665404	664691	R	yebC family-(CT457)
CPn0574	665945	665394	R	
CPn0575	666494	665982	R	YhhY-Amino Group Acetyl Transferase-(CT458)
CPn0576	667543	666494	R	prfB-Peptide Chain Release Factor 2 (natural UGA frame-shift)-(CT459)
CPn0576	667598	667530	R	prfB-(natural UGA frame-shift)
CPn0577	667895	668155	F	SWIB (YM74) complex protein-(CT460)
CPn0578	668406	669365	F	yaeI-phosphohydrolase-(CT461)
CPn0579	669361	669993	F	ygbP/yacM-Sugar Nucleotide Phosphorylase-(CT462)
CPn0580	669993	670793	F	truA-Pseudouridylate Synthase I-(CT463)
CPn0581	671434	670745	R	Phosphoglycolate Phosphatase-(CT464)
CPn0582	671503	672177	F	CT465 hypothetical protein
CPn0583	672400	672717	F	CT466 hypothetical protein
CPn0584	672707	673798	F	atoS/ntrB-2-Component Sensor-(CT467)
CPn0585	675817	673865	F	*similarity to Cps IncA ₂
CPn0586	676026	677183	F	atoC/ntrC-2-Component Regulator-(CT468)
CPn0587	677441	678124	F	*yvyD_Bs conserved hypothetical protein
CPn0588	678084	678626	F	CT469 hypothetical protein
CPn0589	678640	679395	F	CT470 hypothetical protein
CPn0590	680112	679516	F	CT471 hypothetical protein
CPn0591	680373	681020	F	yagE family-(CT472)
CPn0592	681153	681461	F	yidD family-(CT473)
CPn0593	682476	681391	F	CT474 hypothetical protein
CPn0594	682583	684958	F	pheT-phenylalanyl tRNA Synthetase Beta-(CT475)
CPn0595	684958	685926	F	CT476 hypothetical protein
CPn0596	685939	686457	F	ada-methyltransferase-(CT477)
CPn0597	688215	686479	R	oppC ₂ -Oligopeptide Permease ₂ -(CT478)
CPn0598	689697	688219	R	oppB ₂ -Oligopeptide Permease ₂ -(CT479)
CPn0599	691802	689682	R	oppA ₅ -oligopeptide Binding Lipoprotein ₅ -(CT480)
CPn0600	692147	691827	R	
CPn0601	693053	692736	R	CT483 hypothetical protein
CPn0602	694105	693104	R	CT484 hypothetical protein
CPn0603	694205	695185	F	hemZ-Ferrochetalase-(CT485)
CPn0604	695945	695196	R	fliY-Glutamine Binding Protein-(CT486)
CPn0605	696707	696150	R	yhhF-Methylase -(CT487)
CPn0606	697444	696707	R	CT488 hypothetical protein
CPn0607	698895	697573	R	glgC-Glucose-1-P Adenyltransferase-(CT489)
CPn0608	699645	699016	R	*pyrF-Uridine 5'-Monophosphate Synthase (Ump Synthase)-truncated?
CPn0609	699705	699986	F	CT490 hypothetical protein
CPn0610	701420	700029	R	rho-Transcription Termination Factor-(CT491)
CPn0611	702025	701420	R	yacE-predicted phosphatase/kinase-(CT492)
CPn0612	704631	702022	R	polA-DNA Polymerase I-(CT493)
CPn0613	705656	704658	R	sohB-Protease-(CT494)
CPn0614	707402	705783	R	adt ₂ -ADP/ATP Translocase ₂ -(CT495)
CPn0615	708137	707634	R	pgsA ₁ -Glycerol-3-P Phosphatidyltransferase ₁ -(CT496)
CPn0616	708791	710137	F	dnaB-Replicative DNA Helicase-(CT497)
CPn0617	710484	712316	F	gidA-FAD-dependent oxidoreductase-(CT498)
CPn0618	712306	713010	F	lplA-Lipoate-Protein Ligase A-(CT499)
CPn0619	713444	713013	R	ndk-Nucleoside-2-P Kinase-(CT500)
CPn0620	714139	713519	R	ruvA-Holliday Junction Helicase-(CT501)
CPn0621	714647	714144	R	ruvC-Crossover Junction Endonuclease-(CT502)
CPn0622	715752	714793	R	CT503 hypothetical protein
CPn0623	716993	716163	R	CT504 hypothetical protein
CPn0624	718015	717011	R	gapA-Glyceraldehyde-3-P Dehydrogenase-(CT505)
CPn0625	718485	718060	R	rl17-L17 Ribosomal Protein-(CT506)
CPn0626	719616	718495	R	rpoA-RNA Polymerase Alpha-(CT507)
CPn0627	720038	719640	R	rs11-S11 Ribosomal Protein-(CT508)
CPn0628	720428	720063	R	rs13-S13 Ribosomal Protein-(CT509)
CPn0629	721857	720487	R	secY-Translocase-(CT510)
CPn0630	722316	721885	R	rl15-L15 Ribosomal Protein-(CT511)
CPn0631	722806	722312	R	rs5-S5 Ribosomal Protein-(CT512)
CPn0632	723195	722827	R	rl18-L18 Ribosomal Protein-(CT513)
CPn0633	723757	723209	R	rl6-L6 Ribosomal Protein-(CT514)
CPn0634	724185	723787	R	rs8-S8 Ribosomal Protein-(CT515)
CPn0635	724745	724206	R	rl5-L5 Ribosomal Protein-(CT516)
CPn0636	725082	724750	R	rl24-L24 Ribosomal Protein-(CT517)
CPn0637	725464	725099	R	rl14-L14 Ribosomal Protein-(CT518)
CPn0638	725747	725490	R	rs17-S17 Ribosomal Protein-(CT519)

CPn0639	725958	725743	R	rl29-L29 Ribosomal Protein-(CT520)
CPn0640	726377	725964	R	rl16-L16 Ribosomal Protein-(CT521)
CPn0641	727077	726409	R	rs3-S3 Ribosomal Protein-(CT522)
CPn0642	727428	727096	R	rl22-L22 Ribosomal Protein-(CT523)
CPn0643	727713	727450	R	rs19-S19 Ribosomal Protein-(CT524)
CPn0644	728573	727722	R	rl2-L2 Ribosomal Protein-(CT525)
CPn0645	728930	728598	R	rl23-L23 Ribosomal Protein-(CT526)
CPn0646	729621	728950	R	rl4-L4 Ribosomal Protein-(CT527)
CPn0647	730331	729657	R	rl3-L3 Ribosomal Protein-(CT528)
CPn0648	731603	730605	R	CT529 hypothetical protein
CPn0649	732672	731710	R	fmt-Methionyl tRNA Formyltransferase-(CT530)
CPn0650	733501	732665	R	lpxA-Acyl-Carrier UDP-GlcNAc -(CT531)
CPn0651	733975	733517	R	fabZ-Myristoyl-Acyl Carrier Dehydratase-(CT532)
CPn0652	734835	733990	R	lpxC-Myristoyl GlcNAc Deacetylase-(CT533)
CPn0653	736490	734868	R	cutE-Apolipoprotein N-Acetyltransferase-(CT534)
CPn0654	736967	736503	R	vdID/yciA-acyl-CoA Thioesterase-(CT535)
CPn0655	737847	737101	R	dnaQ_2-DNA Pol III Epsilon Chain_2-(CT536)
CPn0656	737872	738048	F	
CPn0657	738473	738051	R	yjeE (ATPase or Kinase)-(CT537)
CPn0658	739168	738455	R	CT538 hypothetical protein
CPn0659	739533	739838	F	trxA-Thioredoxin-(CT539)
CPn0660	740327	739860	R	spoU_2-rRNA Methylase_2-(CT540)
CPn0661	741100	740327	R	mip-FKBP-type peptidyl-prolyl cis-trans isomerase-(CT541)
CPn0662	742923	741172	R	aspS-Aspartyl tRNA Synthetase-(CT542)
CPn0663	744190	742901	R	hisS-Histidyl tRNA Synthetase-(CT543)
CPn0664	744757	744557	R	
CPn0665	745001	746365	F	uhpC-Hexosphosphate Transport -(CT544)
CPn0666	746388	750107	F	dnaE-DNA Pol III Alpha-(CT545)
CPn0667	751058	750177	R	predicted OMP (leader (17))-(CT546)
CPn0668	751209	752162	F	CT547 hypothetical protein
CPn0669	752179	752775	F	CT548 hypothetical protein
CPn0670	752765	753196	F	rsbW-sigma regulatory factor-histidine kinase-(CT549)
CPn0671	753630	753205	R	CT550 hypothetical protein
CPn0672	753741	755048	F	dacF(pbp5)-D-Ala-D-Ala Carboxypeptidase-(CT551)
CPn0673	755287	755463	F	CT552 hypothetical protein
CPn0674	756668	755577	R	fmu-RNA Methyltransferase-(CT553)
CPn0675	757919	756768	R	CT696 hypothetical protein
CPn0676	759217	758051	R	homologous to CT695
CPn0677	760401	759256	R	
CPn0678	761320	760682	R	
CPn0679	762930	761725	R	pgk-Phosphoglycerate Kinase-(CT693)
CPn0680	764248	762971	R	ygo4-Phosphate Permease-(CT692)
CPn0681	764929	764258	R	CT691 hypothetical protein
CPn0682	764984	765955	F	dppD-ABC ATPase Dipeptide Transport-(CT690)
CPn0683	765948	766919	F	dppF-ABC ATPase Dipeptide Transport-(CT689)
CPn0684	768038	767181	R	spoJ/parB-Chromosome Partitioning Protein-(CT688)
CPn0685	768068	768217	F	
CPn0686	768361	768176	R	
CPn0687	768564	769214	F	CT482 hypothetical protein
CPn0688	769382	770137	F	CT481 hypothetical protein
CPn0689	771404	770187	R	yfhO_1-Nifs-related Aminotransferase_1-(CT687)
CPn0690	772680	771436	R	ABC Transporter Membrane Protein-(CT686)
CPn0691	773452	772685	R	abcX-ABC Transporter ATPase-(CT685)
CPn0692	774912	773461	R	ABC Transporter-(CT684)
CPn0693	776256	775240	R	TPR Repeats (O-Linked GlcNAc Transferase similarity)-(CT683)
CPn0694	779599	776330	R	pbp2-PBP2-transglycolase/transpeptidase-(CT682)
CPn0695	780216	781382	F	ompA-Major Outer Membrane Protein-(CT681)
CPn0696	781769	782599	F	rs2-S2 Ribosomal Protein-(CT680)
CPn0697	782602	783447	F	tsf-Elongation Factor TS-(CT679)
CPn0698	783458	784201	F	pyrH-UMP Kinase-(CT679)
CPn0699	784182	784721	F	rrf-Ribosome Releasing Factor-(CT677)
CPn0700	785097	785609	F	CT676 hypothetical protein
CPn0701	785599	786672	F	karG-Arginine Kinase-(CT675)
CPn0702	789685	786929	R	yscC/gspD-Yop C/Gen Secretion Protein D-(CT674)
CPn0703	791190	789685	R	pkn5-S/T Protein Kinase-(CT673)
CPn0704	792321	791209	R	fliN- Flagellar Motor Switch Domain/YscQ family-(CT672)
CPn0705	793173	792334	R	CT671 hypothetical protein
CPn0706	793683	793180	R	CT670 hypothetical protein
CPn0707	795029	793704	R	yscN-Yop N (Flagellar-Type ATPase)-(CT669)
CPn0708	795705	795034	R	CT668 hypothetical protein
CPn0709	796188	795742	R	CT667 hypothetical protein
CPn0710	796461	796210	R	CT666 hypothetical protein

CPn0711	796731	796486	R	CT665 hypothetical protein
CPn0712	799315	796781	R	FHA domain; homology to adenylate cyclase)-(CT664)
CPn0713	799721	799332	R	CT663 hypothetical protein
CPn0714	801107	800091	R	hemA-Glutamyl tRNA Reductase-(CT662)
CPn0715	801657	803462	F	gyrB_2-DNA Gyrase Subunit B_2-(CT661)
CPn0716	803469	804902	F	gyrA_2-DNA Gyrase Subunit A_2-(CT660)
CPn0717	805010	805306	F	CT656 hypothetical protein
CPn0718	805309	805626	F	CT657 hypothetical protein
CPn0719	805916	806890	F	sfhB-(Pseudouridine Synthase)-(CT658)
CPn0720	807003	807236	F	CT659 hypothetical protein
CPn0721	807683	808489	F	kdsA-KDO Synthetase-(CT655)
CPn0722	808489	808974	F	CT654 hypothetical protein
CPn0723	808984	809703	F	yhbG-ABC Transporter ATPase-(CT653)
CPn0724	810527	809706	R	
CPn0725	810811	810587	R	CT652.1 hypothetical protein
CPn0726	813372	810880	R	CT620 hypothetical protein
CPn0727	813577	816192	F	CT619 hypothetical protein
CPn0728	818477	816525	R	CHLPN 76kDa Homolog_1 (CT622)
CPn0729	819857	818592	R	CHLPN 76kDa Homolog_2 (CT623)
CPn0730	821603	819963	R	mviN-Integral Membrane Protein-(CT624)
CPn0731	821587	821760	F	
CPn0732	822098	822976	F	nfo-Endonuclease IV-(CT625)
CPn0733	823727	823101	R	rs4-S4 Ribosomal Protein-(CT626)
CPn0734	823944	824915	F	yceA-(CT627)
CPn0735	825668	825003	R	*pyrH/udk-Uridine Kinase (Uridine Monophosphokinase) (Pyrimidine Ribonucleoside Kinase).
CPn0736	827686	825992	R	ygeD-Efflux Protein-(CT641)
CPn0737	827685	830756	F	recC-Exodeoxyribonuclease V, Gamma-(CT640)
CPn0738	830746	833895	F	recB-Exodeoxyribonuclease V, Beta-(CT639)
CPn0739	834871	833861	R	CT638 hypothetical protein
CPn0740	836048	834864	R	tyrB-Aromatic AA Aminotransferase-(CT637)
CPn0741	838350	836185	R	greA-Transcription Elongation Factor-(CT636)
CPn0742	838463	838888	F	CT635 hypothetical protein
CPn0743	838962	840362	F	nqrA-Ubiquinone Oxidoreductase, Alpha-(CT634)
CPn0744	841384	840389	R	hemB-Porphobilinogen Synthase-(CT633)
CPn0745	841903	841742	R	
CPn0746	841975	843567	F	CT632 hypothetical protein
CPn0747	843675	843740	F	CT631 hypothetical protein
CPn0747	843725	843910	F	CT631 hypothetical protein (frame-shift)
CPn0748	844987	844121	R	ispa-Geranyl Transtransferase-(CT628)
CPn0749	845629	845006	R	glmU-UDP-GlcNAc Pyrophosphorylase-(CT629)
CPn0750	846411	845707	R	tctD/cpxR-HTH Transcriptional Regulatory Protein + Receiver Domain-(CT630)
CPn0751	846608	848434	F	CT651 hypothetical protein
CPn0752	848604	850082	F	recD_2-Exodeoxyribonuclease V, Alpha_2-(CT652)
CPn0753	851006	850161	R	
CPn0754	851336	851040	R	rs20-S20 Ribosomal Protein-(CT617)
CPn0755	851597	852799	F	CT616 hypothetical protein
CPn0756	852961	854676	F	rpoD-RNA Polymerase Sigma-66 -(CT615)
CPn0757	854733	855134	F	folX-Dihydroneopterin Aldolase-(CT614)
CPn0758	855110	856459	F	folP/dhps-Dihydropteroate Synthase-(CT613)
CPn0759	856488	856997	F	folA-Dihydrofolate Reductase-(CT612)
CPn0760	856957	857694	F	CT611 hypothetical protein
CPn0761	857704	858375	F	CT610 hypothetical protein
CPn0762	859597	858539	R	recA-RecA recombination protein-(CT650)
CPn0763	860511	859972	R	ygfA-Formyltetrahydrofolate Cycloligase-(CT649)
CPn0764	861807	860524	R	CT648 hypothetical protein
CPn0765	862382	861801	R	CT647 hypothetical protein
CPn0766	863782	862394	R	CT646 hypothetical protein
CPn0767	863884	864177	F	CT645 hypothetical protein
CPn0768	864159	865163	F	yohI/nir3-predicted oxidoreductase -(CT644)
CPn0769	867733	865121	R	topA-DNA Topoisomerase I-Fused to SWI Domain-(CT643)
CPn0770	868340	869131	F	CT642 hypothetical protein
CPn0771	870463	869144	R	rpoN-RNA Polymerase Sigma-54-(CT609)
CPn0772	872385	870469	R	uvrD-DNA Helicase-(CT608)
CPn0773	872488	873195	F	ung-Uracil DNA Glycosylase-(CT607)
CPn0774	873195	873425	F	CT606.1 hypothetical protein
CPn0775	874031	873414	R	yggV family-(CT606)
CPn0776	874246	875487	F	CT605 hypothetical protein
CPn0777	875601	877178	F	groEL_2-heat shock protein-60 -(CT604)
CPn0778	877505	878092	F	tsa/ahpC-Thio-specific Antioxidant (TSA) Peroxidase-(CT603)
CPn0779	878481	878095	R	CT602 hypothetical protein

CPn0780	879205	878591	R	papQ/amiB-N-Acetylmuramoyl-L-Ala Amidase-(CT601)
CPn0781	879773	879198	R	pal-Peptidoglycan-Associated Lipoprotein-(CT600)
CPn0782	881065	879773	R	tolB-polysaccharide transporter-(CT599)
CPn0783	881885	881100	R	CT598 hypothetical protein
CPn0784	882296	881892	R	exbD-Biopolymer Transport Protein-(CT597)
CPn0785	882991	882296	R	exbB/tolQ-polysaccharide transporter-(CT596)
CPn0786	883185	885293	F	dsbD/xprA-Thio:disulfide Interchange Protein-(CT595)
CPn0787	885619	886401	F	yabD/ycfH-PHP superfamily (urease/pyrimidinase) hydrolase-(CT594)
CPn0788	886542	887432	F	sdhC-Succinate Dehydrogenase-(CT593)
CPn0789	887439	889316	F	sdhA-Succinate Dehydrogenase-(CT592)
CPn0790	889330	890103	F	sdhB-Succinate Dehydrogenase-(CT591)
CPn0791	893050	890111	R	CT590 hypothetical protein
CPn0792	894919	893108	R	CT589 hypothetical protein
CPn0793	896823	894919	R	rbsU-sigma regulatory family protein-PP2C phosphatase (RsbW antagonist)-(CT588)
CPn0794	897174	898004	F	
CPn0795	898128	899195	F	
CPn0796	899301	901340	F	
CPn0797	901600	902694	F	
CPn0798	902846	903856	F	
CPn0799	904986	903940	R	
CPn0800	906532	905249	R	eno-Enolase-(CT587)
CPn0801	908697	906727	R	uvrB-Exonuclease ABC Subunit B-(CT586)
CPn0802	909740	908709	R	trpS-Tryptophanyl tRNA Synthetase-(CT585)
CPn0803	910303	909752	R	CT584 hypothetical protein
CPn0804	911059	910310	R	gp6D-CHLTR Plasmid Paralog-(CT583)
CPn0805	911831	911067	R	minD-chromosome partitioning ATPase-CHLTR plasmid protein GP5D-(CT582)
CPn0806	913771	911867	R	thrS-Threonyl tRNA Synthetase-(CT581)
CPn0807	913971	914879	F	CT580 hypothetical protein
CPn0808	916287	914956	R	CT579 hypothetical protein
CPn0809	917785	916307	R	CT578 hypothetical protein
CPn0810	918184	917825	R	CT577 hypothetical protein
CPn0811	918900	918208	R	lcrH_1-Low Ca Response Protein H_1-(CT576)
CPn0812	919123	920862	F	mutL-DNA Mismatch Repair-(CT575)
CPn0813	920870	921934	F	pepP-Aminopeptidase P-(CT574)
CPn0814	922107	923357	F	CT573 hypothetical protein
CPn0815	923361	925622	F	gspD/pilQ-Gen. Secretion Protein D-(CT572)
CPn0816	925615	927102	F	gspE-Gen. Secretion Protein E-(CT571)
CPn0817	927115	928287	F	gspF-Gen. Secretion Protein F-(CT570)
CPn0818	928314	928682	F	predicted OMP [leader (16) peptide]-(CT569)
CPn0819	928689	929132	F	CT568 hypothetical protein
CPn0820	929120	929659	F	CT567 hypothetical protein
CPn0821	929667	930668	F	CT566 hypothetical protein
CPn0822	930756	931229	F	CT565 hypothetical protein
CPn0823	932367	931501	R	yscT/spaR-YopT Translocation T-(CT564)
CPn0824	932662	932378	R	yscS/fliQ-YopS/fliQ Translocation Protein-(CT563)
CPn0825	933594	932677	R	yscR-Yop Translocation R-(CT562)
CPn0826	934310	933612	R	yscL-Yop Translocation L-(CT561)
CPn0827	935264	934434	R	CT560 hypothetical protein
CPn0828	936271	935267	R	yscJ-Yop Translocation J-(CT559)
CPn0829	936744	937298	F	
CPn0830	937444	937959	F	
CPn0831	938267	938434	F	
CPn0832	939747	938827	R	lipA-Lipoate Synthetase-(CT558)
CPn0833	941129	939747	R	lpdA-Lipoamide Dehydrogenase-(CT557)
CPn0834	941553	942014	F	CT556 hypothetical protein
CPn0835	945689	942045	R	mot1_1-SWI/SNF-family-helicase_1-(CT555)
CPn0836	946879	945722	R	brnQ-Amino Acid (Branched) Transport-(CT554)
CPn0837	947771	947145	R	nth-Endonuclease III-(CT697)
CPn0838	949106	947781	R	thdF-Thiophene/Furan Oxidation Protein-(CT698)
CPn0839	949257	950159	F	psdD-Phosphatidylserine Decarboxylase-(CT699)
CPn0840	950222	951544	F	CT700 hypothetical protein
CPn0841	951731	954640	F	secA_2-Translocase SecA_2-(CT701)
CPn0842	954883	954710	R	CT702 hypothetical protein (frame-shift with 0843)
CPn0843	955191	954994	R	CT702 hypothetical protein
CPn0844	956730	955270	R	yphC-GTPase/GTP-binding protein-(CT703)
CPn0845	958079	956650	R	pcnB_1-Poly A Polymerase_1-(CT704)
CPn0846	959374	958112	R	clpX-CLP Protease ATPase-(CT705)
CPn0847	959995	959387	R	clpP-CLP Protease Subunit-(CT706)
CPn0848	961502	960177	R	tig/murI-Trigger Factor-peptidyl-prolyl isomerase-(CT707)
CPn0849	961788	965285	F	mot1_2-SWI/SNF family helicase_2-(CT708)
CPn0850	965293	966390	F	mreB-Rod Shape Protein-Sugar Kinase-(CT709)

CPn0851	966396	968195	F	pckA-Phosphoenolpyruvate Carboxykinase-(CT710)
CPn0852	968316	970613	F	CT711 hypothetical protein
CPn0853	970637	971803	F	CT712 hypothetical protein
CPn0854	972837	971806	R	ompB-Outer Membrane Protein B-(CT713)
CPn0855	973995	972994	R	gpdA-Glycerol-3-P Dehydrogenase-(CT714)
CPn0856	975377	973995	R	AgX-1 Homolog-UDP-Glucose Pyrophosphorylase-(CT715)
CPn0857	975757	975392	R	CT716 hypothetical protein
CPn0858	977055	975757	R	fliI-Flagellum-specific ATP Synthase-(CT717)
CPn0859	977588	977055	R	CT718 hypothetical protein
CPn0860	978630	977608	R	fliF-Flagellar M-Ring Protein-(CT719)
CPn0861	979722	978925	R	nifU-NifU-related protein-(CT720)
CPn0862	980873	979722	R	yfhO_2-NifS-related protein_2-(CT721)
CPn0863	981514	980831	R	pgmA-Phosphoglycerate Mutase-(CT722)
CPn0864	981670	982374	F	yjbC-predicted pseudouridine synthase-(CT723)
CPn0865	982418	982942	F	CT724 hypothetical protein
CPn0866	983491	982916	R	birA-Biotin Synthetase-(CT725)
CPn0867	983423	984667	F	rodA-Rod Shape Protein-(CT726)
CPn0868	986643	984670	F	zntA/cadA-Metal Transport P-type ATPase-(CT727)
CPn0869	987401	986658	F	CT728 hypothetical protein
CPn0870	988728	987448	F	serS-Seryl tRNA Synthetase_2-(CT729)
CPn0871	988772	989899	F	ribD-Riboflavin Deaminase-(CT730)
CPn0872	989963	991216	F	ribA&ribB-GTP Cyclohydratase & DHBP Synthase -(CT731)
CPn0873	991233	991694	F	ribE-Ribityllumazine Synthase-(CT732)
CPn0874	993107	991749	F	CT733 hypothetical protein
CPn0875	993372	994022	F	CT734 hypothetical protein
CPn0876	994144	995517	F	dagA_2-D-Alanine/Glycine Permease_2-(CT735)
CPn0877	995533	995982	F	ybcL family-(CT736)
CPn0878	996654	995992	F	SET Domain protein-(CT737)
CPn0879	997439	996645	R	yycJ-metal dependent hydrolase-(CT738)
CPn0880	999861	997444	R	ftsK-Cell Division Protein FtsK-(CT739)
CPn0881	1005667	1006209	F	
CPn0882	1006268	1007404	F	
CPn0883	1008865	1007573	R	dmpP/nqr6-Phenolhydrolase/NADH ubiquinone oxidoreductase-(CT740)
CPn0884	1009359	1009009	R	CT741 hypothetical protein
CPn0885	1010635	1009433	R	ygcA-rRNA Methyltransferase-(CT742)
CPn0886	1011276	1010908	R	hctA-Histone-Like Developmental Protein-(CT743)
CPn0887	1011692	1014157	F	CHLTR possible phosphoprotein-(CT744)
CPn0888	1015423	1014119	R	hemG-protoporphyrinogen Oxidase-(CT745)
CPn0889	1016835	1015462	R	hemN_2-Coproporphyrinogen III Oxidase_2-(CT746)
CPn0890	1017805	1016819	R	hemE-Uroporphyrinogen Decarboxylase-(CT747)
CPn0891	1021073	1017819	R	mfd-Transcription-Repair Coupling-(CT748)
CPn0892	1023661	1021046	R	alaS-Alanyl tRNA Synthetase-(CT749)
CPn0893	1023894	1025888	F	tktB-Transketolase-(CT750)
CPn0894	1026766	1025888	R	amn-AMP Nucleosidase-(CT751)
CPn0895	1026988	1027557	F	efp_2-Elongation Factor P_2-(CT752)
CPn0896	1027595	1027822	F	CT753 hypothetical protein
CPn0897	1028737	1027853	R	(possible phosphohydrolase)-(CT754)
CPn0898	1030460	1028904	R	Mitochondrial HSP60 Chaperonin Homolog-(CT755)
CPn0899	1030875	1032215	F	murF-Muramoyl-DAP Ligase-(CT756)
CPn0900	1032235	1033281	F	mraY-Muramoyl-Pentapeptide Transferase-(CT757)
CPn0901	1033287	1034537	F	murD-Muramoylalanine-Glutamate Ligase-(CT758)
CPn0902	1034543	1035241	F	nlpD-Muramidase (invasin repeat family)-(CT759)
CPn0903	1035263	1036417	F	ftsW-Cell Division Protein FtsW-(CT760)
CPn0904	1036326	1037396	F	murG-Peptidoglycan Transferase-(CT761)
CPn0905	1037409	1039835	F	murC&ddlA-Muramate-Ala Ligase & D-Ala-D-Ala Ligase-(CT762)
CPn0906	1040340	1039915	R	CT763 hypothetical protein
CPn0907	1040780	1040445	R	cutA-Periplasmic Divalent Cation Tolerance Protein CutA (C-Type Cytochrome Biogenesis Protein)
CPn0908	1041589	1040780	R	CT764 hypothetical protein
CPn0909	1041637	1041966	F	rsbV_2-Sigma Factor Regulator_2-(CT765)
CPn0910	1041979	1043004	F	miaA-tRNA Pyrophosphate Transferase-(CT766)
CPn0911	1044043	1042985	R	Fe-S cluster oxidoreductase_2-(CT767)
CPn0912	1044129	1045760	F	CT768 hypothetical protein
CPn0913	1045760	1045945	F	
CPn0914	1045999	1046397	F	
CPn0915	1046461	1046817	F	ybeB-iojap superfamily ortholog-(CT769)
CPn0916	1046837	1048084	F	fabF-Acyl Carrier Protein Synthase-(CT770)
CPn0917	1048090	1048539	F	hydrolase/phosphatase homolog-(CT771)
CPn0918	1049223	1048579	R	ppa-Inorganic Pyrophosphatase-(CT772)
CPn0919	1049378	1050430	F	ldh-Leucine Dehydrogenase-(CT773)
CPn0920	1051405	1050431	R	cysQ-Sulfite Synthesis/biphosphate phosphatase-(CT774)
CPn0921	1051535	1052293	F	snGlycerol-3-P Acyltransferase-(CT775)

CPn0922	1052314	1053927	F	aas-Acylglycerophosphoethanolamine Acyltransferase-(CT776)
CPn0923	1053984	1055093	F	bioF_1-Oxononanoate Synthase_1-(CT777)
CPn0924	1057274	1055028	R	priA-Primosomal Protein N' -(CT778)
CPn0925	1057900	1057226	R	CT779 hypothetical protein
CPn0926	1058060	1058557	F	Thioredoxin Disulfide Isomerase-(CT780)
CPn0927	1059809	1058670	R	*CHLPS 43 kDa protein homolog_2
CPn0928	1061008	1059884	R	*CHLPS 43 kDa protein homolog_3
CPn0929	1062292	1061186	R	*CHLPS 43 kDa protein homolog_4
CPn0930	1062857	1063330	F	
CPn0931	1064138	1065718	F	lysS-Lysyl tRNA Synthetase-(CT781)
CPn0932	1067142	1065721	R	cysS-Cysteiny tRNA Synthetase-(CT782)
CPn0933	1067535	1068578	F	predicted disulfide bond isomerase-(CT783)
CPn0934	1068942	1068526	R	rnpA-Ribonuclease P Protein Component-(CT784)
CPn0935	1069091	1068957	R	rl34-L34 Ribosomal Protein-(CT785)
CPn0936	1069336	1069470	F	rl36-L36 Ribosomal Protein-(CT786)
CPn0937	1069496	1069798	F	rs14-S14 Ribosomal Protein-(CT787)
CPn0938	1070322	1069849	R	CT788 hypothetical protein -(leader (60) peptide-periplasmic)
CPn0939	1070728	1071195	F	CT790 hypothetical protein
CPn0940	1073012	1071204	R	uvrC-Excinuclease ABC, Subunit C-(CT791)
CPn0941	1075501	1073018	R	mutS-DNA Mismatch Repair-(CT792)
CPn0942	1075985	1077754	F	dnaG/prim-DNA Primase-(CT794)
CPn0943	1077978	1078238	F	CT794.1 hypothetical protein
CPn0944	1078512	1078997	F	
CPn0945	1079070	1079660	F	CT795 hypothetical protein
CPn0946	1082786	1079745	R	glyQ-Glycyl tRNA Synthetase-(CT796)
CPn0947	1083442	1084059	F	pgsA_2-Glycerol-3-P-Phosphatidyltransferase_2-(CT797)
CPn0948	1085474	1084047	R	glgA-Glycogen Synthase-(CT798)
CPn0949	1085929	1086483	F	ctc-General Stress Protein-(CT799)
CPn0950	1086488	1087027	F	pth-Peptidyl tRNA Hydrolase-(CT800)
CPn0951	1087122	1087457	F	rs6-S6 Ribosomal Protein-(CT801)
CPn0952	1087478	1087723	F	rs18-S18 Ribosomal Protein-(CT802)
CPn0953	1087742	1088248	F	rl9-L9 Ribosomal Protein-(CT803)
CPn0954	1088286	1088708	F	ychB-Predicted Kinase-(CT804)
CPn0955	1088612	1089175	F	(frame-shift with 0954)
CPn0956	1089560	1090909	F	CT805 hypothetical protein
CPn0957	1093788	1090963	R	ide/ptr-Insulinase family/Protease III-(CT806)
CPn0958	1094785	1093793	R	plsB-Glycerol-3-P Acyltransferase-(CT807)
CPn0959	1096343	1094799	R	cafE-Axial Filament Protein-(CT808)
CPn0960	1096764	1097102	F	CT809 hypothetical protein
CPn0961	1097118	1097297	F	rl32-L32 Ribosomal Protein-(CT810)
CPn0962	1097316	1098275	F	plsX-FA/Phospholipid Synthesis Protein-(CT811)
CPn0963	1098398	1103224	F	pmp_21-Polymorphic Outer Membrane Protein D Family-(CT812)
CPn0964	1104758	1103301	R	
CPn0965	1106736	1104925	R	lpxB-Lipid A Disaccharide Synthase-(CT411)
CPn0966	1108037	1106748	R	pcnB_2-PolyA Polymerase_2-(CT410)
CPn0967	1108512	1109885	F	mrsA/pgm-Phosphoglucomutase-(CT815)
CPn0968	1109895	1111721	F	glmS-Glucosamine-Fructose-6-P Aminotransferase-(CT816)
CPn0969	1111812	1112999	F	0969-tyrP_1-Tyrosine Transport_1-(CT817) tyrP_1-Tyrosine Transport_1-(CT817)
CPn0970	1113461	1114648	F	0970-tyrP_2-Tyrosine Transport_2-(CT818) tyrP_2-Tyrosine Transport_2-(CT818)
CPn0971	1114702	1115415	F	yccA-Transport Permease-(CT819)
CPn0972	1116299	1115430	R	ftsY-Cell Division Protein FtsY-(CT820)
CPn0973	1116370	1117527	F	sucC-Succinyl-CoA Synthetase, Beta-(CT821)
CPn0974	1117544	1118422	F	sucD-Succinyl-CoA Synthetase, Alpha-(CT822)
CPn0975	1119104	1119637	F	
CPn0976	1120082	1121185	F	
CPn0977	1121371	1122402	F	
CPn0978	1122665	1123693	F	
CPn0979	1123980	1125443	F	htrA-DO Serine Protease-(CT823)
CPn0980	1126982	1125504	R	*similarity to Saccharomyces cerevisiae hypothetical 52.9KD protein
CPn0981	1127031	1129952	F	Zinc Metalloprotease (insulinase family)-(CT824)
CPn0982	1131194	1129962	R	yigN family-(CT825)
CPn0983	1132000	1131206	R	psaA-Glycerol-Serine Phosphatidyltransferase-(CT826)
CPn0984	1132379	1135510	F	nrdA-Ribonucleoside Reductase, Large Chain-(CT827)
CPn0985	1135534	1136571	F	nrdB-Ribonucleoside Reductase, Small Chain-(CT828)
CPn0986	1136724	1137395	F	yggH-predicted rRNA Methylase-(CT829)
CPn0987	1137516	1138115	F	ytgB-like predicted rRNA methylase-(CT830)
CPn0988	1138986	1138075	R	murB-UDP-N-Acetylenolpyruvoylglucosamine Reductase-(CT831)
CPn0989	1139495	1139016	R	CT832 hypothetical protein
CPn0990	1139883	1140440	F	infC-Initiation Factor 3-(CT833)
CPn0991	1140421	1140612	F	rl35-L35 Ribosomal Protein-(CT834)

CPn0992	1140634	1140996	F	r120-L20 Ribosomal Protein-(CT835)
CPn0993	1141014	1142030	F	pheS-Phenylalanyl tRNA Synthetase, Alpha-(CT836)
CPn0994	1142398	1144440	F	CT837 hypothetical protein
CPn0995	1145512	1144415	R	CT838 hypothetical protein
CPn0996	1146589	1145519	R	CT839 hypothetical protein
CPn0997	1146708	1147664	F	mesJ-PP-loop superfamily ATPase-(CT840)
CPn0998	1147855	1150584	F	ftsH-ATP-dependent zinc protease-(CT841)
CPn0999	1152847	1150766	R	pnp-Polyribonucleotide Nucleotidyltransferase-(CT842)
CPn1000	1153157	1152891	R	rs15-S15 Ribosomal Protein-(CT843)
CPn1001	1153405	1153869	F	yfhC-cytosine deaminase-(CT844)
CPn1002	1153862	1154089	F	CT845 hypothetical protein
CPn1003	1154796	1154092	R	CT846 hypothetical protein
CPn1004	1155397	1154879	R	CT847 hypothetical protein
CPn1005	1155933	1155415	R	CT848 hypothetical protein
CPn1006	1156472	1155990	R	CT849 hypothetical protein
CPn1007	1156689	1156907	F	CT849.1 hypothetical protein
CPn1008	1156928	1158223	F	CT850 hypothetical protein
CPn1009	1159058	1158186	R	map-Methionine Aminopeptidase-(CT851)
CPn1010	1159672	1159067	R	CT852 hypothetical protein
CPn1011	1160306	1159902	R	CT853 hypothetical protein
CPn1012	1162193	1160421	R	yzcB-ABC transporter permease-(CT854)
CPn1013	1162245	1163624	F	fumC-Fumarate Hydratase-(CT855)
CPn1014	1165426	1163732	R	ychM-Sulfate Transporter-(CT856)
CPn1015	1165634	1166893	F	CT857 hypothetical protein (possible IM protein)
CPn1016	1167042	1168898	F	CT858 hypothetical protein
CPn1017	1169006	1169935	F	lytB-Metalloprotease-(CT859)
CPn1018	1169898	1170629	F	
CPn1019	1172128	1170638	R	CT860 hypothetical protein
CPn1020	1173679	1172150	R	CT861 hypothetical protein
CPn1021	1174213	1173698	R	lcrH_2-Low Calcium Response_2-(CT862)
CPn1022	1175673	1174216	R	CT863 hypothetical protein
CPn1023	1176035	1176331	F	
CPn1024	1177236	1176334	R	xerD-Integrase/recombinase-(CT864)
CPn1025	1177302	1178879	F	pgi-Glucose-6-P Isomerase-(CT378)
CPn1026	1178997	1179137	F	ltuA-(CT377)
CPn1027	1179175	1180755	F	
CPn1028	1181016	1181999	F	mdhC-Malate Dehydrogenase-(CT376)
CPn1029	1182008	1182844	F	
CPn1030	1183886	1182843	R	predicted D-amino acid dehydrogenase-(CT375)
CPn1031	1185552	1184098	R	arcD-Arginine/Ornithine Antiporter-(CT374)
CPn1032	1186150	1185566	R	CT373 hypothetical protein
CPn1033	1187500	1186187	R	CT372 hypothetical protein
CPn1034	1188517	1187732	R	Predicted OMP_1 (CT371) [leader (18) peptide]
CPn1035	1190000	1188570	R	AroE-Shikimate 5-Dehydrogenase-(CT370)
CPn1036	1191135	1189984	R	AroB-Dehydroquinase Synthase-(CT369)
CPn1037	1192199	1191123	R	AroC-Chorismate Synthase-(CT368)
CPn1038	1192726	1192199	R	aroL-Shikimate Kinase II-(CT367)
CPn1039	1193999	1192665	R	aroA-Phosphoshikimate Vinyltransferase-(CT366)
CPn1040	1194741	1194073	R	
CPn1041	1195994	1194726	R	*bioA-Adenosylmethionine-8-Amino-7-Oxononanoate Aminotransferase
CPn1042	1196590	1195934	R	*bioD-dethiobiotin synthetase
CPn1043	1197717	1196572	R	bioF_2-Oxononanoate Synthase_2
CPn1044	1198691	1197699	R	*bioB-Biotin Synthase
CPn1045	1199590	1198901	R	*conserved hypothetical bacterial membrane protein
CPn1046	1200675	1199590	R	*Tryptophan Hydroxylase
CPn1047	1200552	1201343	F	dapB-Dihydrodipicolinate Reductase-(CT364)
CPn1048	1201606	1202604	F	asd-Aspartate Dehydrogenase-(CT363)
CPn1049	1202595	1203914	F	lysC-Aspartokinase III-(CT362)
CPn1050	1203926	1204798	F	dapA-Dihydrodipicolinate Synthase-(CT361)
CPn1051	1204962	1205270	F	
CPn1052	1205417	1206169	F	
CPn1053	1206153	1206701	F	
CPn1054	1207034	1209466	F	
CPn1055	1209694	1210521	F	
CPn1056	1210527	1211228	F	
CPn1057	1211497	1213596	F	CT356 hypothetical protein
CPn1058	1213748	1214836	F	CT355 hypothetical protein
CPn1059	1214848	1215678	F	kgsA-Dimethyladenosine Transferase-(CT354)
CPn1060	1217658	1215727	R	dxs/ckt-Transketolase-(CT331)
CPn1061	1217920	1217666	R	CT330 hypothetical protein
CPn1062	1219820	1218159	R	xseA-Exodoxyribonuclease VII-(CT329)
CPn1063	1219951	1220712	F	tpiS-Triosephosphate Isomerase-(CT328)

Table 2 (Supplemental Data) Functional Assignments of *C. pneumoniae* Coding Sequences. *C. trachomatis* genes are shown in parentheses.

5			Amino Acid Biosynthesis
		<i>Aromatic Family</i>	
	1039 (CT366)	aroA	Phosphoshikimate Vinyltransferase
	1036 (CT369)	aroB	Dehydroquinate Synthase
	1037 (CT368)	aroC	Chorismate Synthase
10	1035 (CT370)	aroE	Shikimate 5-Dehydrogenase
	0484 (CT382)	aroG	Deoxyheptonate Aldolase
	1038 (CT367)	aroL	Shikimate Kinase II
	0740 (CT637)	tyrB	Aromatic AA Aminotransferase
		<i>Aspartate Family (lysine)</i>	
15	1048 (CT363)	asd	Aspartate Dehydrogenase
	1050 (CT361)	dapA	Dihydrodipicolinate Synthase
	1047 (CT364)	dapB	Dihydrodipicolinate Reductase
	0519 (CT430)	dapF	Diaminopimelate Epimerase
	1049 (CT362)	lysC	Aspartokinase III
20		<i>Serine Family</i>	
	0433 (CT282)	gcsH	Glycine Cleavage System H Protein
	0521 (CT432)	glyA	Serine Hydroxymethyltransferase
		<i>Base & Nucleotide Metabolism</i>	
	0171	guaA	GMP Synthase
25	0172	guaB	Inosine 5'-Monophosphate Dehydrogenase
	0608		Uridine 5'-Monophosphate Synthase
	0735		Uridine Kinase
	0244 (CT128)	adk	Adenylate Kinase
	0894 (CT751)	amn	AMP Nucleosidase
30	0568 (CT452)	cmk	CMP Kinase
	0392 (CT039)	dcd	dCTP Deaminase
	0059 (CT292)	dut	dUTP Nucleotidohydrolase
	0120 (CT030)	gmK	GMP Kinase
	0619 (CT500)	ndk	Nucleoside-2-P Kinase
35	0984 (CT827)	nrdA	Ribonucleoside Reductase, Large Chain
	0985 (CT828)	nrdB	Ribonucleoside Reductase, Small Chain
	0236 (CT183)	pyrG	CTP Synthetase
	0698 (CT678)	pyrH	UMP Kinase
	0273 (CT188)	tdk	Thymidylate Kinase
40	0659 (CT539)	trxA	Thioredoxin
	0314 (CT099)	trxB	Thioredoxin Reductase
	1001 (CT844)	yfhC	Cytosine Deaminase
			Biosynthesis of Cofactors
45		<i>Biotin, Lipoate & Ubiquinone</i>	
	1041	bioA	Adenosylmethionine-8-Amino-7-Oxononanoate Aminotransferase
	1044	bioB	Biotin Synthase
	1042	bioD	Dethiobiotin Synthetase
	0923 (CT777)	bioF_1	Oxononanoate Synthase_1
50	1043 (CT777)	bioF_2	Oxononanoate Synthase_2
	0866 (CT725)	birA	Biotin Synthetase
	0748 (CT628)	ispA	Geranyl Transtransferase
	0832 (CT558)	lipA	Lipoate Synthetase

	0265	(CT219)	ubiA	Benzoate Octaphenyltransferase
	0264	(CT220)	ubiD	Phenylacrylate Decarboxylase
	0515	(CT428)	ubiE	Ubiquinone Methyltransferase
	<i>Folic Acid</i>			
5	0759	(CT612)	folA	Dihydrofolate Reductase
	0335	(CT078)	folD	Methylene Tetrahydrofolate Dehydrogenase
	0758	(CT613)	folP	Dihydropteroate Synthase
	0757	(CT614)	folX	Dihydroneopterin Aldolase
	0763	(CT649)	ygfA	Formyltetrahydrofolate Cycloligase
10	<i>Porphyrin</i>			
	0714	(CT662)	hemA	Glutamyl tRNA Reductase
	0744	(CT633)	hemB	Porphobilinogen Synthase
	0052	(CT299)	hemC	Porphobilinogen Deaminase
	0890	(CT747)	hemE	Uroporphyrinogen Decarboxylase
15	0888	(CT745)	hemG	protoporphyrinogen Oxidase
	0138	(CT210)	hemL	Glutamate-1-Semialdehyde-2,1-Aminomutase
	0380	(CT052)	hemN_1	Coproporphyrinogen III Oxidase_1
	0889	(CT746)	hemN_2	Coproporphyrinogen III Oxidase_2
	0603	(CT485)	hemZ	Ferrochetalase
20	<i>Riboflavin</i>			
	0872	(CT731)	ribA&ribB	GTP Cyclohydrtase & DHBP Synthase
	0532	(CT405)	ribC	Riboflavin Synthase
	0871	(CT730)	ribD	Riboflavin Deaminase
	0873	(CT732)	ribE	Ribityllumazine Synthase
25	0320	(CT093)	ribF	FAD Synthase
	<i>Cell Envelope</i>			
	<i>Fatty Acid & Phospholipid Metabolism</i>			
	0161	(CT206)		(predicted acyltransferase family)
30	0922	(CT776)	aas	Acylglycerophosphoethanolamine Acyltransferase
	0414	(CT265)	accA	AcCoA Carboxylase/Transferase Alpha
	0183	(CT123)	accB	Biotin Carboxyl Carrier Protein
	0182	(CT124)	accC	Biotin Carboxylase
	0058	(CT293)	accD	AcCoA Carboxylase/Transferase Beta
35	0295	(CT236)	acpP	Acyl Carrier Protein
	0313	(CT100)	acpS	Acyl-carrier Protein Synthase
	0567	(CT451)	cdsA	Phosphatidate Cytidylyltransferase
	0297	(CT238)	fabD	Malonyl Acyl Carrier Transcyclase
	0916	(CT770)	fabF	Acyl Carrier Protein Synthase
40	0296	(CT237)	fabG	Oxoacyl (Carrier Protein) Reductase
	0298	(CT239)	fabH	Oxoacyl Carrier Protein Synthase III
	0406	(CT104)	fabI	Enoyl-Acyl-Carrier Protein Reductase
	0651	(CT532)	fabZ	Myristoyl-Acyl Carrier Dehydratase
	0098	(CT010)	htrB	Acyltransferase
45	0271	(CT136)		Lysophospholipase Esterase
	0615	(CT496)	pgsA_1	Glycerol-3-P Phosphatidyltransferase_1
	0947	(CT797)	pgsA_2	Glycerol-3-P Phosphatidyltransferase_2
	0958	(CT807)	plsB	Glycerol-3-P-Acyltransferase
	0569	(CT453)	plsC	Glycerol-3-P Acyltransferase
50	0962	(CT811)	plsX	FA/Phospholipid Synthesis Protein
	0839	(CT699)	psdD	Phosphatidylserine Decarboxylase
	0983	(CT826)	pssA	Glycerol-Serine Phosphatidyltransferase
	0921	(CT775)		snGlycerol-3-P Acyltransferase
	0654	(CT535)	yciA	Acyl-CoA Thioesterase
55	0877	(CT736)	ybcL	CT736 Hypothetical Protein

LPS

	0154	(CT208)	gseA	KDO Transferase
	0721	(CT655)	kdsA	KDO Synthetase
	0235	(CT182)	kdsB	Deoxyoctulosonic Acid Synthetase
5	0650	(CT531)	lpxA	Acyl-Carrier UDP-GlcNAc O-Acyltransferase
	0965	(CT411)	lpxB	Lipid A Disaccharide Synthase
	0652	(CT533)	lpxC	Myristoyl GlcNAc Deacetylase
	0302	(CT243)	lpxD	UDP Glucosamine N-Acyltransferase

Membrane Proteins, Lipoproteins & Porins

	0310	(CT251)	60IM	60kDa Inner Membrane Protein
10	0556	(CT442)	crpA	15kDa Cysteine-Rich Protein
	0653	(CT534)	cutE	Apolipoprotein N-Acetyltransferase
	0311	(CT252)	lgt	Protein Lipoprotein Diacylglycerol Transferase
	0558	(CT444)	omcA	9kDa-Cysteine-Rich Lipoprotein
	0557	(CT443)	omcB	60kDa Cysteine-Rich OMP
15	0695	(CT681)	ompA	Major Outer Membrane Protein
	0854	(CT713)	ompB	Outer Membrane Protein B
	0781	(CT600)	pal	Peptidoglycan-Associated Lipoprotein
	0300	(CT241)	yaeT	Omp85 Homolog

Peptidoglycan

20	0417	(CT268)	amiA	N-Acetylmuramoyl Alanine Amidase
	0780	(CT601)	amiB	N-Acetylmuramoyl-L-Ala Amidase
	0672	(CT551)	dacF	D-Ala-D-Ala Carboxypeptidase
	0968	(CT816)	glmS	Glucosamine-Fructose-6-P Aminotransferase
	0749	(CT629)	glmU	UDP-GlcNAc Pyrophosphorylase
25	0900	(CT757)	mraY	Muramoyl-Pentapeptide Transferase
	0571	(CT455)	murA	UDP-N-Acetylglucosamine Transferase
	0988	(CT831)	murB	UDP-N-Acetylenolpyruvoylglucosamine Reductase
	0905	(CT762)	murC&ddlA	Muramate-Ala Ligase & D-Ala-D-Ala Ligase
	0901	(CT758)	murD	Muramoylalanine-Glutamate Ligase
30	0418	(CT269)	murE	N-Acetylmuramoylalanine-glutamyl DAP Ligase
	0899	(CT756)	murF	Muramoyl-DAP Ligase
	0904	(CT761)	murG	Peptidoglycan Transferase
	0902	(CT759)	nlpD	Muramidase (invasin repeat family)
	0694	(CT682)	pbp2	PBP2-Transglycosylase/Transpeptidase
35	0419	(CT270)	pbp3	Transglycosylase/Transpeptidase
	0421	(CT272)	yabC	PBP2B Family Methyltransferase

Cellular Processes

Cell Division

40	0959	(CT808)	cafE	Axial Filament Protein
	0880	(CT739)	ftsK	Cell Division Protein FtsK
	0903	(CT760)	ftsW	Cell Division Protein FtsW
	0972	(CT820)	ftsY	Cell Division Protein FtsY
	0617	(CT498)	gidA	FAD-dependent Oxidoreductase
45	0805	(CT582)	minD	Chromosome Partitioning ATPase
	0850	(CT709)	mreB	Rod Shape Protein-Sugar Kinase
	0867	(CT726)	rodA	Rod Shape Protein
	0684	(CT688)	parB	Chromosome-Partitioning Protein

Detoxification

50	0057	(CT294)	sodM	Superoxide Dismutase (Mn)
	0778	(CT603)	ahpC	Thio-specific Antioxidant (TSA) Peroxidase

Signal Transduction

	0148	(CT145)		S/T Protein Kinase
	0584	(CT467)	atoS	Two-Component Sensor
55	0294	(CT235)		cAMP-Dependent Protein Kinase Regulatory Subunit
	0712	(CT664)		(FHA domain)

	0478	(CT379)	hflX	GTP Binding Protein
	0703	(CT673)		S/T Protein Kinase
	0095	(CT301)		S/T Protein Kinase
	0397	(CT259)		PP2C Phosphatase Family
5	0037	(CT337)	ptsH	PTS Phosphocarrier Protein Hpr
	0038	(CT336)	ptsI	PTS PEP Phosphotransferase
	0060	(CT291)	ptsN_1	PTS IIA Protein_1
	0061	(CT290)	ptsN_2	PTS IIA Protein + HTH DNA-Binding Domain
	0262	(CT218)	surE	SurE-like Acid Phosphatase
10	0838	(CT698)	thdF	Thiophene/Furan Oxidation Protein
	0693	(CT683)		TPR Repeats-CT683 Hypothetical Protein
	0321	(CT092)	ychF	GTP Binding Protein
	0544	(CT418)	yhbZ	GTP binding protein
	0844	(CT703)	yphC	GTPase/GTP-binding protein
15	<i>Standard Protein Secretion</i>			
	0115	(CT025)	fliH	Signal Recognition Particle GTPase
	0363	(CT060)	fliA	Flagellar Secretion Protein
	0858	(CT717)	fliI	Flagellum-specific ATP Synthase
	0704	(CT672)	fliN	Flagellar Motor Switch Domain/YscQ family
20	0815	(CT572)	gspD	Gen. Secretion Protein D
	0816	(CT571)	gspE	Gen. Secretion Protein E
	0817	(CT570)	gspF	Gen. Secretion Protein F
	0359	(CT064)	lepA	GTPase
	0110	(CT020)	lepB	Signal Peptidase I
25	0535	(CT408)	lspA	Lipoprotein Signal Peptidase
	0260	(CT141)	secA_1	Protein Translocase Subunit_1
	0841	(CT701)	secA_2	Translocase SecA_2
	0564	(CT448)	secD&secF	Protein Export Proteins SecD/SecF (fusion)
30	0075	(CT321)	secE	Preprotein Translocase
	0629	(CT510)	secY	Translocase
	0848	(CT707)	tig	Trigger Factor-Peptidyl-prolyl Isomerase
	<i>Transport-Related Proteins</i>			
	0486			Hypothetical Proline Permease
	0289	(CT230)	aaaT	Neutral Amino Acid (Glutamate) Transporter
35	0691	(CT685)	abcX	ABC Transporter ATPase
	1031	(CT374)	arcD	Arginine/Ornithine Antiporter
	0482	(CT381)	artJ	Arginine Periplasmic Binding Protein
	0836	(CT554)	brnQ	Amino Acid (Branched) Transport
	0536	(CT409)	dagA_1	D-Ala/Gly Permease_1
40	0876	(CT735)	dagA_2	D-Alanine/Glycine Permease_2
	0682	(CT690)	dppD	ABC ATPase Dipeptide Transport
	0683	(CT689)	dppF	ABC ATPase Dipeptide Transport
	0280	(CT689)	dppF	Dipeptide Transporter ATPase
	0785	(CT596)	exbB	Macromolecule Transporter
45	0784	(CT597)	exbD	Biopolymer Transport Protein
	0604	(CT486)	fliY	Glutamine Binding Protein
	0192	(CT129)	glnP	ABC Amino Acid Transporter Permease
	0191	(CT130)	glnQ	ABC Amino Acid Transporter ATPase
	0528	(CT401)	gltT	Glutamate Symport
50	0286	(CT194)	mgtE	Mg ⁺⁺ Transporter (CBS Domain)
	0413	(CT264)	msbA	Transport ATP Binding Protein
	0290	(CT231)		Na ⁺ -dependent Transporter
	0195	(CT198)	oppA_1	Oligopeptide Binding Protein_1
	0196	(CT198)	oppA_2	Oligopeptide Binding Protein_2
55	0197	(CT139)	oppA_3	Oligopeptide Binding Protein_3
	0198	(CT175)	oppA_4	Oligopeptide Binding Protein_4

0599	(CT480)	oppA_5	Oligopeptide Binding Lipoprotein_5
0199	(CT199)	oppB_1	Oligopeptide Permease_1
0598	(CT479)	oppB_2	Oligopeptide Permease_2
0200	(CT200)	oppC_1	Oligopeptide Permease_1
0597	(CT478)	oppC_2	Oligopeptide Permease_2
0201	(CT201)	oppD	Oligopeptide Transport ATPase
0202	(CT202)	oppF	Oligopeptide Transport ATPase
0231	(CT180)	tauB	ABC Transport ATPase (Nitrate/Fe)
0782	(CT599)	tolB	Macromolecule Transporter
0969	(CT817)	tyrP_1	Tyrosine Transport_1
0970	(CT818)	tyrP_2	Tyrosine Transport_2
0665	(CT544)	uhpC	Hexosphosphate Transport
0282	(CT216)	xasA	Amino Acid Transporter
0207	(CT204)	ybhI	dicarboxylate Translocator
0971	(CT819)	yccA	Transport Permease
0248	(CT152)	ycfV	ABC Transporter ATPase
1014	(CT856)	ychM	Sulfate Transporter
0736	(CT641)	ygeD	Efflux Protein
0680	(CT692)	ygo4	Phosphate Permease
0723	(CT653)	yhbG	ABC Transporter ATPase
0023	(CT348)	yjiK	ABC Transporter Protein ATPase
0127	(CT034)	ytfF	Cationic Amino Acid Transporter
0349	(CT067)	ytgA	Solute Protein Binding Family
0348	(CT068)	ytgB	ABC transporter ATPase
0347	(CT069)	ytgC	Integral Membrane Protein
0346	(CT070)	ytgD	Integral Membrane Protein
1012	(CT854)	yzcB	ABC Transporter Permease
0868	(CT727)	zntA	Metal Transport P-type ATPase
0279			Possible ABC Transporter Permease Protein
0543	(CT417)		(Metal Transport Protein)
0692	(CT684)		ABC Transporter
0542	(CT416)		ABC Transporter ATPase
0690	(CT686)		ABC Transporter Membrane Protein
0541	(CT415)		solute binding protein
<i>Type-III Secretion</i>			
0323	(CT090)	lcrD	Low Calcium Response D
0324	(CT089)	lcrE	Low Calcium Response E
0811	(CT576)	lcrH_1	Low Ca Response Protein H_1
1021	(CT862)	lcrH_2	Low Calcium Response_2
0325	(CT088)	sycE	Secretion Chaperone
0702	(CT674)	yscC	Yop C/Gen Secretion Protein D
0828	(CT559)	yscJ	Yop Translocation J
0826	(CT561)	yscL	Yop Translocation L
0707	(CT669)	yscN	Yop N (Flagellar-Type ATPase)
0825	(CT562)	yscR	Yop Translocation R
0824	(CT563)	yscS	YopS Translocation Protein
0823	(CT564)	yscT	YopT Translocation T
0322	(CT091)	yscU	Yop Translocation Protein U

50

Central Intermediary Metabolism

Glycogen Metabolism

0856	(CT715)		UDP-Glucose Pyrophosphorylase
0948	(CT798)	glgA	Glycogen Synthase
0475	(CT866)	glgB	Glucan Branching Enzyme
0607	(CT489)	glgC	Glucose-1-P Adenyltransferase
0307	(CT248)	glgP	Glycogen Phosphorylase
0388	(CT042)	glgX	Glycogen Hydrolase (debranching)

	0360	(CT063)	gnd	6-Phosphogluconate Dehydrogenase
	0185	(CT121)	rpe	Ribulose-P Epimerase
	0141	(CT213)	rpiA	Ribose-5-P Isomerase A
	0083	(CT313)	tal	Transaldolase
5	0893	(CT750)	tktB	Transketolase
	0238	(CT185)	zwf	Glucose-6-P Dehydrogenase
	<i>Pyruvate Dehydrogenase</i>			
	0833	(CT557)	lpdA	Lipoamide Dehydrogenase
10	0436	(CT285)	lplA_1	Lipoate Protein Ligase-Like Protein
	0618	(CT499)	lplA_2	Lipoate-Protein Ligase A
	0033	(CT340)	pdhA&B	Oxoisovalerate Dehydrogenase α/β Fusion
	0304	(CT245)	pdhA	Pyruvate Dehydrogenase Alpha
	0305	(CT246)	pdhB	Pyruvate Dehydrogenase Beta
	0306	(CT247)	pdhC	Dihydrolipoamide Acetyltransferase
15	<i>TCA Cycle</i>			
	0495	(CT390)	aspC	Aspartate Aminotransferase
	1013	(CT855)	fumC	Fumarate Hydratase
	1028	(CT376)	mdhC	Malate Dehydrogenase
	0789	(CT592)	sdhA	Succinate Dehydrogenase
20	0790	(CT591)	sdhB	Succinate Dehydrogenase
	0788	(CT593)	sdhC	Succinate Dehydrogenase
	0378	(CT054)	sucA	Oxoglutarate Dehydrogenase
	0377	(CT055)	sucB_1	Dihydrolipoamide Succinyltransferase_1
	0527	(CT400)	sucB_2	Dihydrolipoamide Succinyltransferase_2
25	0973	(CT821)	sucC	Succinyl-CoA Synthetase, Beta
	0974	(CT822)	sucD	Succinyl-CoA Synthetase, Alpha
	Protein Folding, Assembly & Modification			
	<i>Chaperones</i>			
30	0949	(CT799)	ctc	General Stress Protein
	0534	(CT407)	dksA	DnaK Suppressor
	0032	(CT341)	dnaJ	Heat Shock Protein J
	0503	(CT396)	dnaK	Hsp-70
	0134	(CT110)	groEL_1	Hsp-60_1
35	0777	(CT604)	groEL_2	Hsp-60_2
	0898	(CT755)	groEL_3	Hsp-60_3
	0135	(CT111)	groES	10KDa Chaperonin
	0502	(CT395)	grpE	HSP-70 Cofactor
	0661	(CT541)	mip	FKBP-type Peptidyl-prolyl Cis-Trans Isomerase
40	<i>Proteases</i>			
	0144	(CT113)	clpB	Clp Protease ATPase
	0437	(CT286)	clpC	ClpC Protease
	0520	(CT431)	clpP_1	CLP Protease
	0847	(CT706)	clpP_2	CLP Protease Subunit
45	0846	(CT705)	clpX	CLP Protease ATPase
	0269	(CT138)		Dipeptidase
	0998	(CT841)	ftsH	ATP-dependent Zinc Protease
	0030	(CT343)	gcp_1	O-Sialoglycoprotein Endopeptidase_1
	0194	(CT197)	gcp_2	O-Sialoglycoprotein Endopeptidase_2
50	0979	(CT823)	htrA	DO Serine Protease
	0957	(CT806)	ide	Insulinase family/Protease III
	0027	(CT344)	lon	Lon ATP-dependent Protease
	1017	(CT859)	lytB	Metalloprotease
	1009	(CT851)	map	Methionine Aminopeptidase
55	0385	(CT045)	pepA	Leucyl Aminopeptidase A
	0136	(CT112)	pepF	Oligopeptidase

	0813	(CT574)	pepP	Aminopeptidase P
	0613	(CT494)	sohB	Protease
	0555	(CT441)	tsp	Tail-Specific Protease
	0344	(CT072)	yaeL	Metalloprotease
5	0981	(CT824)		Zinc Metalloprotease (insulinase family)

Protein Isomerases

	0227	(CT176)	dsbB	Disulfide bond Oxidoreductase
	0786	(CT595)	dsbD	Thio:disulfide Interchange Protein
	0228	(CT177)	dsbG	Disulfide Bond Chaperone
10	0933	(CT783)		Predicted Disulfide Bond Isomerase
	0926	(CT780)		Thioredoxin Disulfide Isomerase

00431353124100

	0932	(CT782)	cysS	CysteinyI tRNA Synthetase
	0003	(CT003)	gatA	Glu tRNA Gln Amidotransferase (A subunit)
	0004	(CT004)	gatB	Glu tRNA Gln Amidotransferase (B Subunit)
	0002	(CT002)	gatC	Glu tRNA Gln Amidotransferase (C subunit)
5	0560	(CT445)	gltX	Glutamyl-tRNA Synthetase
	0946	(CT796)	glyQ	Glycyl tRNA Synthetase
	0663	(CT543)	hisS	Histidyl tRNA Synthetase
	0109	(CT019)	ileS	Isoleucyl-tRNA Synthetase
	0153	(CT209)	leuS	Leucyl tRNA Synthetase
10	0931	(CT781)	lysS	Lysyl tRNA Synthetase
	0122	(CT032)	metG	Methionyl-tRNA Synthetase
	0993	(CT836)	pheS	Phenylalanyl tRNA Synthetase, Alpha
	0594	(CT475)	pheT	Phenylalanyl tRNA Synthetase Beta
	0500	(CT393)	proS	Prolyl tRNA Synthetase
15	0870	(CT729)	serS	Seryl tRNA Synthetase_2
	0806	(CT581)	thrS	Threonyl tRNA Synthetase
	0802	(CT585)	trpS	Tryptophanyl tRNA Synthetase
	0361	(CT062)	tyrS	Tyrosyl tRNA Synthetase
	0094	(CT302)	valS	Valyl tRNA Synthetase
20	<i>Peptide Chain Initiation, Elongation & Termination</i>			
	1067	(CT353)	def	Polypeptide Deformylase
	0184	(CT122)	efp_1	Elongation Factor P_1
	0895	(CT752)	efp_2	Elongation Factor P_2
	0550	(CT437)	fusA	Elongation Factor G
25	0073	(CT323)	infA	Initiation Factor IF-1
	0317	(CT096)	infB	Initiation Factor-2
	0990	(CT833)	infC	Initiation Factor 3
	0113	(CT023)	pfrA	Peptide Chain Releasing Factor 1
	0576	(CT459)	prfB	Peptide Chain Release Factor 2
30	0950	(CT800)	pth	Peptidyl tRNA Hydrolase
	0318	(CT095)	rbfA	Ribosome Binding Factor A
	0699	(CT677)	rrf	Ribosome Releasing Factor
	0697	(CT679)	tsf	Elongation Factor TS
	0074	(CT322)	tufA	Elongation Factor Tu
35	<i>Ribosomal Proteins</i>			
	0078	(CT318)	r11	L1 Ribosomal Protein
	0644	(CT525)	r12	L2 Ribosomal Protein
	0647	(CT528)	r13	L3 Ribosomal Protein
	0646	(CT527)	r14	L4 Ribosomal Protein
40	0635	(CT516)	r15	L5 Ribosomal Protein
	0633	(CT514)	r16	L6 Ribosomal Protein
	0080	(CT316)	r17	L7/L12 Ribosomal Protein
	0953	(CT803)	r19	L9 Ribosomal Protein
	0079	(CT317)	r110	L10 Ribosomal Protein
45	0077	(CT319)	r111	L11 Ribosomal Protein
	0247	(CT125)	r113	L13 Ribosomal Protein
	0637	(CT518)	r114	L14 Ribosomal Protein
	0630	(CT511)	r115	L15 Ribosomal Protein
	0640	(CT521)	r116	L16 Ribosomal Protein
50	0625	(CT506)	r117	L17 Ribosomal Protein
	0632	(CT513)	r118	L18 Ribosomal Protein
	0118	(CT028)	r119	L19 Ribosomal Protein
	0992	(CT835)	r120	L20 Ribosomal Protein
	0546	(CT420)	r121	L21 Ribosomal Protein
55	0642	(CT523)	r122	L22 Ribosomal Protein
	0645	(CT526)	r123	L23 Ribosomal Protein

5	0636	(CT517)	rl24	L24 Ribosomal Protein
	0545	(CT419)	rl27	L27 ribosomal protein
	0327	(CT086)	rl28	L28 Ribosomal Protein
	0639	(CT520)	rl29	L29 Ribosomal Protein
	0112	(CT022)	rl31	L31 Ribosomal Protein
10	0961	(CT810)	rl32	L32 Ribosomal Protein
	0250	(CT150)	rl33	L33 Ribosomal Protein
	0935	(CT785)	rl34	L34 Ribosomal Protein
	0991	(CT834)	rl35	L35 Ribosomal Protein
	0936	(CT786)	rl36	L36 Ribosomal Protein
15	0315	(CT098)	rs1	S1 Ribosomal Protein
	0696	(CT680)	rs2	S2 Ribosomal Protein
	0641	(CT522)	rs3	S3 Ribosomal Protein
	0733	(CT626)	rs4	S4 Ribosomal Protein
	0631	(CT512)	rs5	S5 Ribosomal Protein
20	0951	(CT801)	rs6	S6 Ribosomal Protein
	0551	(CT438)	rs7	S7 Ribosomal Protein
	0634	(CT515)	rs8	S8 Ribosomal Protein
	0246	(CT126)	rs9	S9 Ribosomal Protein
	0549	(CT436)	rs10	S10 Ribosomal Protein
25	0627	(CT508)	rs11	S11 Ribosomal Protein
	0552	(CT439)	rs12	S12 Ribosomal Protein
	0628	(CT509)	rs13	S13 Ribosomal Protein
	0937	(CT787)	rs14	S14 Ribosomal Protein
	1000	(CT843)	rs15	S15 Ribosomal Protein
30	0116	(CT026)	rs16	S16 Ribosomal Protein
	0638	(CT519)	rs17	S17 Ribosomal Protein
	0952	(CT802)	rs18	S18 Ribosomal Protein
	0643	(CT524)	rs19	S19 Ribosomal Protein
	0754	(CT617)	rs20	S20 Ribosomal Protein
35	0031	(CT342)	rs21	S21 Ribosomal Protein
Other Categories				
<i>Chlamydia-Specific Proteins</i>				
40	0561	(CT446)	Euo	CHLPS Euo Protein
	0804	(CT583)	Gp6D	CHLTR Plasmid Paralog
	0186	(CT119)		Similarity to IncA_1
	0291	(CT232)	incB	Inclusion Membrane Protein B
	0292	(CT233)	incC	Inclusion Membrane Protein C
45	1026	(CT377)		LtuA Protein
	0333	(CT080)		LtuB Protein
	0005	(CT871)	pmp_1	Polymorphic Outer Membrane Protein G Family
	0013	(CT871)	pmp_2	Polymorphic Outer Membrane Protein G Family
	0014	(CT871)	pmp_3	Polymorphic Outer Membrane Protein G Family
50	0015	(CT871)	pmp_3	PMP_3 (frame-shift with 0014)
	0016	(CT874)	pmp_4	Polymorphic Outer Membrane Protein G Family
	0017	(CT871)	pmp_4	PMP_4 (frame-shift with 0016)
	0018	(CT874)	pmp_5	Polymorphic Outer Membrane Protein G Family
	0019	(CT871)	pmp_5	PMP_5 (frame-shift with 0018)
55	0444	(CT871)	pmp_6	Polymorphic Outer Membrane Protein G/I Family
	0445	(CT871)	pmp_7	Polymorphic Outer Membrane Protein G Family
	0446	(CT871)	pmp_8	Polymorphic Outer Membrane Protein G Family
	0447	(CT871)	pmp_9	Polymorphic Outer Membrane Protein G/I Family
	0450	(CT871)	pmp_10	Polymorphic Outer Membrane Protein G Family
	0449	(CT871)	pmp_10	PMP_10 (Frame-shift with 0450)

0451	(CT871)	pmp_11	Polymorphic Outer Membrane Protein G Family
0452	(CT874)	pmp_12	Polymorphic Outer Membrane Protein (truncated) A/I Family
0453	(CT871)	pmp_13	Polymorphic Outer Membrane Protein G Family
0454	(CT872)	pmp_14	Polymorphic Outer Membrane Protein H Family
0466	(CT869)	pmp_15	Polymorphic Outer Membrane Protein E Family
0467	(CT869)	pmp_16	Polymorphic Outer Membrane Protein E Family
0468	(CT869)	pmp_17	Polymorphic Outer Membrane Protein E Family
0469	(CT869)	pmp_17	PMP_17 (Frame-shift with 0468)
0470	(CT869)	pmp_17	PMP_17 (Frame-shift with 0469)
0471	(CT870)	pmp_18	Polymorphic Outer Membrane Protein E/F Family
0539	(CT412)	pmp_19	Polymorphic Membrane Protein A Family
0540	(CT413)	pmp_20	Polymorphic Membrane Protein B Family
0963	(CT812)	pmp_21	Polymorphic Membrane Protein D Family
0562			CHLPS 43 kDa Protein Homolog_1
0927			CHLPS 43 kDa Protein Homolog_2
0928			CHLPS 43 kDa Protein Homolog_3
0929			CHLPS 43 kDa Protein Homolog_4
0728	(CT622)		CHLPS 76kDa Homolog_1 (CT622)
0729	(CT623)		CHLPS 76kDa Homolog_2 (CT623)
0133	(CT109)		CHLPS Hypothetical Protein
0332	(CT081)		CHLPS T2 Protein
<i>Miscellaneous Enzymes/Conserved Proteins</i>			
0193		argR	Possible Arginine Repressor
1046			Aromatic Amino Acid Hydroxylase
0232			Similarity to 5'-Methylthioadenosine Nucleosidase
0128	(CT035)		Biotin Protein Ligase
0513	(CT426)		Fe-S Oxidoreductase_1
0911	(CT767)		Fe-S Oxidoreductase_2
0373	(CT057)	gcpE	GcpE Protein
0407	(CT103)		HAD Superfamily Hydrolase/Phosphatase
0917	(CT771)		Hydrolase/Phosphatase Homolog
0488	(CT385)	ycfF	HIT Family Hydrolase
0701	(CT675)	karG	Arginine Kinase
0526	(CT399)	kpsF	GutQ/KpsF Family Sugar-P Isomerase
0919	(CT773)	ldh	Leucine Dehydrogenase
0022	(CT349)	maf	Maf protein
0997	(CT840)	mesJ	PP-loop superfamily ATPase
0151	(CT148)	mhpA	Monooxygenase
0730	(CT624)	mviN	Integral Membrane Protein
0861	(CT720)		NifU-Related Protein
0479	(CT380)	phnP	Metal Dependent Hydrolase
0106	(CT015)	phoH	ATPase
0329	(CT084)		Phospholipase D Superfamily
0435	(CT284)		Phospholipase D Superfamily
0581	(CT464)		Phosphoglycolate Phosphatase
0897	(CT754)		Predicted Phosphohydrolase
0509	(CT422)		Predicted Metalloenzyme
1030	(CT375)		Predicted D-Amino Acid Dehydrogenase
0531	(CT404)		SAM-Dependent Methyltransferase
0337	(CT076)	smpB	Small Protein B
0394	(CT256)	tlyC_1	CBS Domain Protein (Hemolysin Homolog)_1
0510	(CT423)	tlyC_2	CBS Domains (Hemolysin Homolog)_2
0382	(CT048)	yabC	SAM-Dependent Methyltransferase
0787	(CT594)	yabD	PHP Superfamily (Urease/Pyrimidinase) Hydrolase
0611	(CT492)	yacE	Predicted Phosphatase/Kinase
0579	(CT462)	yacM	Sugar Nucleotide Phosphorylase
0578	(CT461)	yaeI	Phosphohydrolase

	0345	(CT071)	yaeM	CT071 Hypothetical Protein
	0566	(CT450)	yaeS	YaeS family Hypothetical Protein
	0591	(CT472)	yagE	YagE family
5	0039	(CT335)	ybaB	YbaB family Hypothetical Protein
	0101	(CT012)	ybbP	YbbP family Hypothetical Protein
	0915	(CT769)	ybeB	iojap Superfamily Ortholog
	0137	(CT108)	ybgI	ACR family
	0529	(CT402)	ycaH	ATPase
10	0438	(CT287)	ycbF	PP-loop Superfamily ATPase
	0734	(CT627)	yceA	YceA Hypothetical Protein
	0954	(CT804)	ychB	Predicted Kinase
	0261	(CT217)	ydaO	PP-Loop Superfamily ATPase
	0245	(CT127)	ydH	Polysaccharide Hydrolase-Invasin Repeat Family
15	0573	(CT457)	yebC	YebC Family Hypothetical Protein
	0689	(CT687)	yfhO_1	NifS-related Aminotransferase_1
	0862	(CT721)	yfhO_2	NifS-related Aminotransferase_2
	0547	(CT434)	ygbB	YgbB Family Hypothetical Protein
	0237	(CT184)	yggF	YggF Family Hypothetical Protein
	0775	(CT606)	yggV	YggV Family Hypothetical Protein
20	0396	(CT258)	yhfO_3	NifS-related Aminotransferase_3
	0605	(CT487)	yhhF	Predicted Methylase
	0575	(CT458)	yhhY	Amino Group Acetyl Transferase
	0592	(CT473)	yidD	YidD Family
25	0982	(CT825)	yigN	YigN Family Hypothetical Protein
	0657	(CT537)	yjeE	YjeE Hypothetical Protein
	0768	(CT644)	yohI	YohI Predicted Oxidoreductase
	0336	(CT077)	yojL	YojL Hypothetical Protein
	0217	(CT140)	ypdP	YpdP Hypothetical Protein
30	0140	(CT212)	yqde	Yqde Hypothetical Protein
	0263	(CT221)	yqfU	YqfU Hypothetical Protein
	0139	(CT211)	yqgE	YqgE Hypothetical Protein
	0270	(CT137)	ywiC	SuA5 Superfamily-related Protein
	0879	(CT738)	yycJ	Metal Dependent Hydrolase
35	Homologs to CHLTR Hypothetical Coding Genes			
	0001	(CT001)	CT001	Hypothetical Protein
	0020	(CT351)	CT351	Hypothetical Protein
	0021	(CT350)	CT350	Hypothetical Protein
40	0026	(CT345)	CT345	Hypothetical Protein
	0035	(CT339)	CT339	Hypothetical Protein
	0036	(CT338)	CT338	Hypothetical Protein
	0055	(CT296)	CT296	Hypothetical Protein
	0062	(CT289)	CT289	Hypothetical Protein
45	0065	(CT288)	CT288	Hypothetical Protein
	0068	(CT360)	CT360	Hypothetical Protein
	0071	(CT325)	CT325	Hypothetical Protein
	0072	(CT324)	CT324	Hypothetical Protein
	0085	(CT311)	CT311	Hypothetical Protein
50	0087	(CT309)	CT309	Hypothetical Protein
	0093	(CT303)	CT303	Hypothetical Protein
	0100	(CT011)	CT011	Hypothetical Protein
	0104	(CT017)	CT017	Hypothetical Protein
	0105	(CT016)	CT016	Hypothetical Protein
55	0107	(CT058)	CT058	Hypothetical Protein_1
	0108	(CT018)	CT018	Similarity
	0111	(CT021)	CT021	Hypothetical Protein
	0121	(CT031)	CT031	Hypothetical Protein

	0129	(CT036)	CT036 Similarity
	0145	(CT114)	CT114 Hypothetical Protein
	0150	(CT147)	CT147 Hypothetical Protein
5	0152	(CT149)	CT149 Hypothetical Protein
	0176	(CT153)	CT153 Hypothetical Protein
	0188	(CT132)	CT132 Hypothetical Protein
	0189	(CT131)	CT131 Hypothetical Protein
	0206	(CT203)	CT203 Hypothetical Protein
10	0229	(CT178)	CT178 Hypothetical Protein
	0230	(CT179)	CT179 Hypothetical Protein
	0234	(CT181)	CT181 Hypothetical Protein
	0249	(CT151)	CT151 Hypothetical Protein
	0253	(CT144)	CT144 Hypothetical Protein_1
15	0254	(CT143)	CT143 Hypothetical Protein_1
	0255	(CT142)	CT142 Hypothetical Protein_1
	0256	(CT144)	CT144 Hypothetical Protein_2
	0257	(CT143)	CT143 Hypothetical Protein_2
	0259	(CT142)	CT142 Hypothetical Protein_2
20	0276	(CT191)	CT191 Hypothetical Protein
	0288	(CT195)	CT195 Hypothetical Protein
	0293	(CT234)	CT234 Hypothetical Protein
	0301	(CT242)	CT368 Hypothetical Protein
	0303	(CT244)	CT244 Hypothetical Protein
	0308	(CT249)	CT249 Similarity
25	0312	(CT101)	CT101 Hypothetical Protein
	0328	(CT085)	CT085 Hypothetical Protein
	0330	(CT083)	CT083 Hypothetical Protein
	0331	(CT082)	CT082 Hypothetical Protein
	0334	(CT079)	CT079 Similarity
30	0342	(CT073)	CT073 Hypothetical Protein
	0343	(CT073)	(frame-shift with 0342?)
	0350	(CT066)	CT066 Hypothetical Protein
	0369	(CT058)	CT058 Hypothetical Protein_2
	0370	(CT058)	CT058 Hypothetical Protein_3
35	0374	(CT056)	CT056 Hypothetical Protein
	0379	(CT053)	CT053 Hypothetical Protein
	0381	(CT326)	CT326 Similarity
	0383	(CT047)	CT047 Hypothetical Protein
	0387	(CT043)	CT043 Hypothetical Protein
40	0389	(CT041)	CT041 Hypothetical Protein
	0393	(CT038)	CT038 Hypothetical Protein
	0395	(CT257)	CT257 Hypothetical Protein
	0399	(CT253)	CT253 Hypothetical Protein
	0400	(CT254)	CT254 Hypothetical Protein
45	0401	(CT255)	CT255 Hypothetical Protein
	0405	(CT105)	CT105 Hypothetical Protein
	0408	(CT102)	CT102 Hypothetical Protein
	0409	(CT260)	CT260 Hypothetical Protein
50	0411	(CT262)	CT262 Hypothetical Protein
	0412	(CT263)	CT263 Hypothetical Protein
	0415	(CT266)	CT266 Hypothetical Protein
	0420	(CT271)	CT271 Hypothetical Protein
	0422	(CT273)	CT273 Hypothetical Protein
	0423	(CT274)	CT274 Hypothetical Protein
55	0425	(CT276)	CT276 Hypothetical Proteins
	0426	(CT277)	CT277 Similarity
	0434	(CT283)	CT283 Hypothetical Protein

	0441	(CT007)	CT007 Hypothetical Protein
	0442	(CT006)	CT006 Hypothetical Protein
	0443	(CT005)	CT005 Hypothetical Protein
	0474	(CT365)	CT365 Hypothetical Protein
5	0476	(CT865)	CT865 Hypothetical Protein
	0480	(CT383)	CT383 Hypothetical Protein
	0485	(CT382)	CT382.1 Hypothetical Protein
	0487	(CT384)	CT384 Hypothetical Protein
	0489	(CT386)	CT386 Hypothetical Protein
10	0490	(CT387)	CT387 Hypothetical Protein
	0491	(CT389)	CT389 Hypothetical Protein
	0496	(CT391)	CT391 Hypothetical Protein
	0497	(CT388)	CT388 Hypothetical Protein
	0506	(CT421)	CT421 Hypothetical Protein
15	0507	(CT421)	CT421.1 Hypothetical Protein
	0508	(CT421)	CT421.2 Hypothetical Protein
	0512	(CT425)	CT425 Hypothetical Protein
	0514	(CT427)	CT427 Hypothetical Protein
	0518	(CT429)	CT429 Hypothetical Protein
20	0522	(CT433)	CT433 Hypothetical Protein
	0525	(CT398)	CT398 Hypothetical Protein
	0533	(CT406)	CT406 Hypothetical Protein
	0537	(CT814)	CT814.1 Hypothetical Protein
	0538	(CT814)	CT814 Hypothetical Protein
25	0554	(CT440)	CT440 Hypothetical Protein
	0559	(CT441)	CT441.1 Hypothetical Protein
	0565	(CT449)	CT449 Hypothetical Protein
	0572	(CT456)	CT456 Hypothetical Protein
	0582	(CT465)	CT465 Hypothetical Protein
30	0583	(CT466)	CT466 Hypothetical Protein
	0588	(CT469)	CT469 Hypothetical Protein
	0589	(CT470)	CT470 Hypothetical Protein
	0590	(CT471)	CT471 Hypothetical Protein
	0593	(CT474)	CT474 Hypothetical Protein
35	0595	(CT476)	CT476 Hypothetical Protein
	0601	(CT483)	CT483 Hypothetical Protein
	0602	(CT484)	CT484 Hypothetical Protein
	0606	(CT488)	CT488 Hypothetical Protein
	0609	(CT490)	CT490 Hypothetical Protein
40	0622	(CT503)	CT503 Hypothetical Protein
	0623	(CT504)	CT504 Hypothetical Protein
	0648	(CT529)	CT529 Hypothetical Protein
	0658	(CT538)	CT538 Hypothetical Protein
	0667	(CT546)	CT546 Hypothetical Protein
45	0668	(CT547)	CT547 Hypothetical Protein
	0669	(CT548)	CT548 Hypothetical Protein
	0671	(CT550)	CT550 Hypothetical Protein
	0673	(CT552)	CT552 Hypothetical Protein
	0675	(CT696)	CT696 Hypothetical Protein
50	0676	(CT695)	CT695 Similarity
	0681	(CT691)	CT691 Hypothetical Protein
	0687	(CT482)	CT482 Hypothetical Protein
	0688	(CT481)	CT481 Hypothetical Protein
	0700	(CT676)	CT676 Hypothetical Protein
55	0705	(CT671)	CT671 Hypothetical Protein
	0706	(CT670)	CT670 Hypothetical Protein
	0708	(CT668)	CT668 Hypothetical Protein

	0709	(CT667)	CT667 Hypothetical Protein
	0710	(CT666)	CT666 Hypothetical Protein
	0711	(CT665)	CT665 Hypothetical Protein
	0713	(CT663)	CT663 Hypothetical Protein
5	0717	(CT656)	CT656 Hypothetical Protein
	0718	(CT657)	CT657 Hypothetical Protein
	0720	(CT659)	CT659 Hypothetical Protein
	0722	(CT654)	CT654 Hypothetical Protein
	0725	(CT652)	CT652.1 Hypothetical Protein
10	0726	(CT620)	CT620 Hypothetical Protein
	0727	(CT619)	CT619 Hypothetical Protein
	0739	(CT638)	CT368 Hypothetical Protein
	0742	(CT635)	CT635 Hypothetical Protein
	0746	(CT632)	CT632 Hypothetical Protein
15	0747	(CT631)	CT631 Hypothetical Protein
	0751	(CT651)	CT651 Hypothetical Protein
	0755	(CT616)	CT616 Hypothetical Protein
	0760	(CT611)	CT611 Hypothetical Protein
	0761	(CT610)	CT610 Hypothetical Protein
20	0764	(CT648)	CT648 Hypothetical Protein
	0765	(CT647)	CT647 Hypothetical Protein
	0766	(CT646)	CT646 Hypothetical Protein
	0767	(CT645)	CT645 Hypothetical Protein
	0770	(CT642)	CT642 Hypothetical Protein
25	0774	(CT606)	CT606.1 Hypothetical Protein
	0776	(CT605)	CT605 Hypothetical Protein
	0779	(CT602)	CT602 Hypothetical Protein
	0783	(CT598)	CT598 Hypothetical Protein
	0791	(CT590)	CT590 Hypothetical Protein
30	0792	(CT589)	CT589 Hypothetical Protein
	0803	(CT584)	CT584 Hypothetical Protein
	0807	(CT580)	CT580 Hypothetical Protein
	0808	(CT579)	CT579 Hypothetical Protein
	0809	(CT578)	CT578 Hypothetical Protein
35	0810	(CT577)	CT577 Hypothetical Protein
	0814	(CT573)	CT573 Hypothetical Protein
	0818	(CT569)	CT569 Hypothetical Protein
	0819	(CT568)	CT568 Hypothetical Protein
	0820	(CT567)	CT567 Hypothetical Protein
40	0821	(CT566)	CT566 Hypothetical Protein
	0822	(CT565)	CT565 Hypothetical Protein
	0827	(CT560)	CT560 Hypothetical Protein
	0834	(CT556)	CT556 Hypothetical Protein
	0840	(CT700)	CT700 Hypothetical Protein
45	0842	(CT702)	CT702 Hypothetical Protein
	0843	(CT702)	CT702 Hypothetical Protein
	0852	(CT711)	CT711 Hypothetical Protein
	0853	(CT712)	CT712 Hypothetical Protein
	0857	(CT716)	CT716 Hypothetical Protein
50	0859	(CT718)	CT718 Hypothetical Protein
	0865	(CT724)	CT724 Hypothetical Protein
	0869	(CT728)	CT728 Hypothetical Protein
	0874	(CT733)	CT733 Hypothetical Protein
	0875	(CT734)	CT734 Hypothetical Protein
55	0884	(CT741)	CT741 Hypothetical Protein
	0887	(CT744)	CHLTR Possible Phosphoprotein
	0896	(CT753)	CT753 Hypothetical Protein

	0980	Similar to <i>Saccharomyces cerevisiae</i> 52.9KDa Protein	
	0232	Similarity to 5'-Methylthioadenosine Nucleosidase	
	1046	Tryptophan Hydroxylase	
5	0477	yqeV_Bs	Conserved Hypothetical Protein
	0048	yqfF-Bs	Conserved Hypothetical IM Protein
	0587	yvyD_Bs	Conserved Hypothetical Protein
	0143	yxjG_Bs_1	Conserved Hypothetical Protein
	0448	yxjG_Bs_2	Conserved Hypothetical Protein
10	0006	0180	0440 0977
	0007	0181	0455 0978
	0008	0190	0456 1018
	0009	0203	0457 1023
	0010	0204	0458 1027
	0011	0205	0459 1029
15	0012	0209	0460 1040
	0028	0210	0461 1051
	0029	0211	0462 1052
	0034	0212	0463 1053
	0041	0213	0464 1054
20	0042	0214	0465 1055
	0043	0215	0472 1056
	0044	0216	0473 1064
	0045	0218	0481 1065
	0046	0220	0483 1066
25	0047	0221	0492 1070
	0049	0222	0493 1071
	0050	0223	0494 1072
	0051	0224	0498
	0063	0225	0499
30	0064	0226	0516
	0066	0233	0517
	0067	0240	0523
	0069	0241	0524
	0070	0242	0553
35	0099	0243	0574
	0124	0266	0600
	0125	0267	0656
	0126	0268	0664
	0130	0277	0677
40	0131	0283	0678
	0132	0284	0685
	0142	0285	0686
	0146	0287	0724
	0147	0352	0731
45	0155	0353	0745
	0156	0354	0753
	0157	0355	0794
	0158	0356	0795
<hr/>			
50	0159	0357	0796
	0162	0358	0797
	0163	0365	0798
	0164	0366	0799
	0165	0367	0829
	0166	0368	0830
55	0167	0371	0831
	0168	0372	0881
	0169	0375	0882

5

0170	0376	0913
0173	0391	0914
0174	0398	0930
0175	0404	0944
0177	0431	0964
0178	0432	0975
0179	0439	0976

004335 4400
0044 999900

As A2

Table 3

Chlamydia pneumoniae Genome Encoded Protein

Cpn_0001 330 4
CT001 hypothetical protein
KRLKDEIKYTLRRKAMLCGIIRGLSSILVLCALNVLIGITHNKNLI IAKLCGGVSTP
ATQITTYIIIGIAGVICLLSFCPCSKSRHSHGDCSSGCHSHSDKI

Cpn_0002 570 875
gata-Glu tRNA Gln Amidotransferase (B Subunit)
IMHSHLPEELTALGAFATVJGSEELDEYATLDAVITIMHFAALSHADGCEGLF
MHVNVNEDLREDSVTSDFNREEFLRNVPESLGGVLRVPAVVK

Cpn_0003 889 2370
gata-Glu tRNA Gln Amidotransferase
KIMYRYSALEAKAVTLGELATATGTHFHRIIEAEGGVGAFISLCKEALQEALIDK
KRSGEPGLKAGVPGVIGIKDNIHVTLGLTKTCAVRLENYQPPFDATVVERIKEDGIILG
KLNMDFEAMGSTTLYSAFHPTNFWLDSRVPGSSGSAASVAFPCVAGSDTGGSRIR
QPAACCGVGVFKPSYGAISRYGLVAFASLQIGPLANTVEDVALMDVFGSRDPKDATS
REFFRDSFMSKLSTEVPKVGIVPRTFLEGLRDDIRENFSSLAIFEGEGTHLVDELIDIL
SHAVSIYIILASAEATNLARFDGVRGYRSPQAHITISQYDLSRGEGFGKVEIRILIG
NVLISAERQNVYKATAVRAKIVKAFRTAFKCEKILAMPVCSAPAFEGIEGIDPVTLYL
QDITVTAMNLAYLPAIAPVSGFSKEGLPLGLQIGQQQGDQVQCVQSGFQEAHQIKQLF
SKRYAKSVVLGGQS

Cpn_0004 2334 3833
gata-B (Pet112) Glu tRNA Gln Amidotransferase (B Subunit)
EICQCCSRSSIMSAYADWSEVIGLEVHVELATASKPLSSALNRFQDEPNINISTVCTG
LPGSLPVLNQSAVEKAVLPGCAVGEISLTSRDRKSYFDPSPNFQITQFEHPII
RKAIVQGEERYFELAQTHIEDDAGMLKHFGFAGVDYNRAGVPLIEIVSKPCMFCEPA
VAYATLSVLSDYIGISDCNMEEGSRFDVNVSVRPGKSPENKVEIKQNSAFMAQA
LEAEQKQIDYELNQPNKDPKLVIPAAVYRWDPKKTVMRLKESAEADYKIFPEPDLPT
LQTSYIERIKRTLPELPHYKYHRYIQEYGLSEDIASLISDKNIATFEVACKDNF
RSLSNWVTFEGGRCKTLGVKLPGSSGIFPEGVAGLVNAIDQGVITGKIAKEIADLMKPS
GKNPEELKEKPELLMSDEGELQKIIAEVLANPESIVDYNGKTKALGFLVGQIMKRT
AGKAPKRVNELLLELDKG

Cpn_0005 4097 6892
pmp-1-Polymorphic Outer Membrane Protein
SDIHFDLQTKMRFSLGFPPLVFTLLSVFDTLSATTISLTPEDSFHGDQNAERSYNY
QAGDVSLTGADVISNVDNLSALNCAFNVTSGSVTFAGNHGLVFNINISSTGKEGAVLC
GAPQATATGVSQFSTLSIFQSPGDIKGLKXNALMLNLYVVRFGQSKTKGGAIS
GANTIVGNYSVSFYQNAATFGCAITHSSGCLQIAVQNAERFAQNTAKNGSGALYSDG
DIDEDQNAIVLFRENEALTGAIGAVCCCLPTSGSTFPIVITFSDNKQLVFERNHISG
GGAATYARKLISISSGGPTLFNNISYANSONLGAIAIDTGGIEISLSAEKGTITFCQNR
SLPLANGIHLQNAKFLKQLNAGSYIEFYDPTSEADGSKTQNLINGDPKMEYTGITLIF
SGERKSLANDPRDKSTIPQNVNLSAGYVIRKEGAETVYSKFTQSPGSHLLDLQTLKIAS
KEDJATGLAIDISLSSSSAAVIRKANTANKQISYVDSIELISPTGNAEYEDLMRNSTQ
PPLSLEPGAGGSVTVTAGDFLVPVPHYGFGQWKLAWITGTGNKVGEEFWDKINYPKPE
KEGILVNIILWGNADVRSLSMQVQETHASSLQTDRLGWDIGINFFHVSASEDNIRYRN
SGGVYLVNNEITPKHYTSMASFQSLSRDKDYAVSNNEYRMYLGSYLQYITSLGNI
ASRNPNVNVGLISRRFLQNPMLIFHPLCAVGHATNDMKTDIYANFPMVNSWRNWCALTEC
GSGMPLLVFENGRLFGALIPMKQLVYAYAGDFKETTADGRFNSGSLTISVPLGRIF
EKLASQDVLYDFSFYIPDIFRDKPSCAAELVISGDSWLVAHVSRHAFVSGTGRYH
FNDYTELLCRGSIIECRPHARYNNINCGSKFRF

Cpn_0006 7299 7141
No robust homolog present in Genebank/EMBL as of 11/7/98
KQLEPLRSALLERLSEMLVLLGVSPETTRSTPEKDANQLPKDSNRNLTSL

Cpn_0007 7488 10496
No robust homolog present in Genebank/EMBL as of 11/7/98
KSERVNLISLFLVAVPLTDSSTSSLSLDEGNQPSMRKRLILAVLIALSILIAQ
GVLLTVATPGLSVISSPAGMGACALGCMVLAIGLDVLLKKEVPIVLSVTTTPOTCS
PRSGEISISGADSTIRSLPTYLLDEGHPQSMRKLRLILAVLIVFSIILASGVLLTVAP
GLSGVSSPAGMGACALGCMVLAIGLDVLLKKEVPIVLSVTTTPOTCSPRSGISIGA
DSTLSTPTPLDEGHPQSMRKLRLILAVLIVFSIILASGVLLTVATPGLSVISSP
AGMGALGCMVLAIGLDVLLKKEVPIVLSVTTTPOTCSPRSGISIGA
PEEMSAFEGYIKVSHENMLKSLPYDGHLEKTKHQRIRVRSLLKAMVPEFLDTRIF
EEEFPPFLSARKRLIDLATLVERKILTEQLERNLNRKAFSYLYQDSIFKKIINDFKLA
WKMILSKSICRFTIIFENHEHGVAKSLHKNVALLKVIYRSLOKSYRDIOMSAKKMI
LHGNPFFSLEDNKKTIMKEHMLSLSSYRKVFLALSDENVDTSPDPKMWLSPICR
DALSEISRDQWQKAHLKHQESLYTQARDRLTQSSKENQKLEKAEQYSSWERYVK
FEIERVQERIRAIQKLYPNILEEETTGQETVPTVQGTASSDLTDILQIEVSSRD
NQSGECSVKVLRSHVEVMSWEVQKQYGPKKKEFQDQMSLERFTEHIELEVLQKDYK
HLSYFKVNNKKEVQYKFRKVLDESQLELAQTESAESLTOEELPLATRGALAKAV
FKGSLCCALAKAKPYFEEDPRDQSDTLRALTLRLQAKASLEEEIFRFSNLENDIAE
ERRLLEKSGQTFERAGLVREIAYESTVDRSLTNTWECTPESEKVFMSYLYNNEEK
RRAKTRLVEMTQRYRDFKMALEAMQNEEALIQEELSIOAPSE

Cpn_0008 10780 11695
No robust homolog present in Genebank/EMBL as of 11/7/98
CKYSYLLNPPPPRRSLGVCSKLSLSITLLVGLVLLTGLTGLTAGISFGAGLGPSA
LQGVLYISGLLLVLRREVPTVRSEIIPRGVSVTSEEPALKAQKEPETKILDLRLPE
LDQDQTYIQEVFACLERLKDPKYEDRGLLTKAEKLRVFDVLEKDMSEFLDQIRVLANE
AYVYEHQCDPLENIAYEIFFSQELRDYACGVGCLYPSGDARADLRKRSVKVEMDRFVR
WLKSWASVMDHSGVARELFFKAVGVLEESVYKLFKSYRDAFYCEKAKIQRDGRFK

Cpn_0009 11689 13119
No robust homolog present in Genebank/EMBL as of 11/7/98
LPSAHARQFRDINXWEDLQKTIFVGGHIDCTDEYVRKSCMWLDROYADKFLREKEE
MERHLEIATMVRKAGIAYAKAAAFKEKRNENRQVQKVKLSGLAEFRNQSSRR
AERLRLATLYPEVSGEVLRLQRTCKVNLNENYADIEKKYHICVREQEHQWKEVENK
PATEYREKELVIAEVEVSFCLELDEKTLKAEESVFEMKFDATKCNKVLSD
VNRLELAEADAEEMIFRTEIEMTLRMVLELPLFMKNTFEKASLQYNSCKEMLAKVEPQ
CGESPTLYVRKRLERLNDQIATYNTQERLAFESDLESKVRICRDLHREOMKHFVEQ
INFINEELIWAALFIDQARIDVATVYMEYLYLQVHIIKREKVRQWAKTERTYREIRO
AFVVMKRLIADDTILKEEDYWLDDWLLRDERKNRORRIICNIAAQQVAVKGF

Cpn_0010 13124 14125
No robust homolog present in Genebank/EMBL as of 11/7/98

CKYFYLRSPYPPPPHSSVGS...
FGLGIGLSALGQVLMISGLL...
PKLEDQDQTDIQEVFACLERLKD...
ALQECIDHKVAVAFDRNSYAMA...
HITWLRDDAKGCCAEKKGMP...
...
HITWLRDDAKGCCAEKKGMP...
...
HITWLRDDAKGCCAEKKGMP...
...

Cpn_0010.1 14268 15746
gata-B (Pet112) Glu tRNA Gln Amidotransferase (B Subunit)
KAPDANKHPPYFFFAV...
FLKRLHRCALAKTTTEKK...
PVEVYSIRENKIQETRSN...
SPEELSSLEQFDHGLKNF...
KEMPCRIEIEKTLMAEL...
RLVSLDEDLRRAYTECQ...
SRLNSEDCTCVSGVKEA...
AKLSEEDKLVKEEYWL...
DQARFLNFRNEDHS

Cpn_0011 15877 16614
gata-B (Pet112) Glu tRNA Gln Amidotransferase (B Subunit)
FWYIMTAAPAILHVSPT...
GLSALIVCGIGISTISL...
QSTKSELEKIEGSRYS...
RLLEFTAANLFKQGRD...
ELEFFF

Cpn_0012 16596 18212
gata-B (Pet112) Glu tRNA Gln Amidotransferase (B Subunit)
GIRVFLPNKYGLKGM...
DKEKAEKAEKAFIEQ...
ERWHMKVLQKLEDDY...
RLDRLARYGCTVDPK...
ARTQLEEQDIAKRE...
DMKARLKLIEDITS...
VFENNYIYVEYVOL...
GFNFMKEKFKAAAK...
AKKLCCLRLDEKEL...
...
AKKLCCLRLDEKEL...
...

Cpn_0013 18509 21106
pmp-3-Polymorphic Outer Membrane Protein
LDRPLAFFIYLLYWK...
LRLDTSTTSFSSKT...
GFEFTFENIDATTAS...
LINDVLIQDNFTSG...
LTLFQNTAAPTAAAG...
KITLRAQAQHTIYF...
KDEKNTSKLLQNV...
NLEINISLNRGKKI...
AGKDIVISADSR...
LLWGSFIDVRSQ...
SYMFGQDITSLG...
ELLPLVSKTLPC...
AEQYVHVAMYS...
LFGNFGEWGRSS...
...

Cpn_0014 21365 21922
pmp-3-Polymorphic Outer Membrane Protein
IQNOSTIYTMKSSFP...
TTLVFKGNVLENI...
VVKSTTIFGSSLS...
PQKLFH

Cpn_0015 21835 24174
pmp-3-Polymorphic Outer Membrane Protein
LEFDKQVSLFSK...
KNGQEVSPSNDT...
TKSTDTVTLTGN...
IAIEDSGELSLAD...
PITGTSSTTVTV...
GGTLLSKHGVTL...
TKATSKNLTLSG...
FHYVGQGTWGP...
ETANEGLOGDR...
LFGDRDLYFVA...
HTNDRLDKTYT...
QFKEQTEAREF...
IRSGDSWKTFT...
...

Cpn_0016 24383 26188
pmp-4-Polymorphic Outer Membrane Protein
RSDFAKRCCHMR...
AATSDASGTTYI...
GVVSVNTAASCI...
GAINTKVLSLGT...
CAGCNLVISNNQ...
GAIYTKKVLSS...
TGCPASVTRNA...
GYIVFZGKLEB...
GRTTFAGAGV...
NLQGVVFLLE...
...

Cpn_0017 26194 27170
pmp-4-Polymorphic Outer Membrane Protein
LIPITWIKETRI...
LGLMURSTSLR...
FAPVLYVYKRI...
HTNEMTTPY...
HDKPRVTHR...
...

ESN-PTS TIA Protein
KLLPEEYEVLVILEKAMPSTYQGNQDLSFLSLLSPRLVMFLGKHRSDETLQDITDILVDA
GLLEDEKQAFDLDALVRRENIMSTGIGMFWAIPHGKLESCNFFFAIGIITQILHWAIIDK
LVRLVFLIGCPENAKAEYIKLLSTLTLSLRRESRROQILQVNTIEVMNVFVGM

CpN_0061 75501 76208
p6n-PT3: IIA Protein + HTH DNA-Binding
RSHECIGKDVMDLKLDEVASLQDVZHTVLQWLKEG... SMNNEYRFGREEIENMLL
HNOALMIOER:EDKEALKDLGLSKYSLYKAHRCGLVCLDVVHSEALQYASKYIAQKQFQ
LDGSLVLEML:HIENLMTGTCIGBIALPHAKDFLINAYDVIIVVPMFLAEIEYCALDGKPK
VGLIFLFAQODKCHLNLVNVKIVHGLMSLNARSFFKNYPKQDQLLAYVKEWESQTH

CpN_0062 76251 77690
CT288 hypothetical protein
KPNP...
LAVVKKPKGSEKQAKKEPRARKGLVPSSTRTLSARAJMKMNSRRKESGQCNIEISANST
PRSVKLRNNKRAEQKAAKQGFSAFNLTLKSLPLPSKQKTSIHEREKATSRFVNSQSL
SSARKRYCTPSSAAPSFLFETIIVRAPVERTKELQDNEIHIPVVOVTNPKEQNTKTQK
LASQASIQOEGTEQSLRELAQASLPVLRNPEVSVORQKEELLKELVAERROCKRKS
VRQALEARSLTKKVARGGSTVTLRYDPEKAAEIKSRNCKVSPAREQKYSSCKRDARA
NGKQDKTTPSEDASQEEQTAGLVKRTPKSQVASNAGNFYRNKNTNIDSYLTANQVSC
SSEETDWPCCSSCVSKRRTHNSISVCTMVTVIAMIVGALIIANATESQTTSDPTPTPTP

CpN_0063 78109 78267
No robust homolog present in Genebank/EMBL as of 11/7/98
PMYANCKHNCCLLYDFSRHSRPPGLPTFTTPPYSTLIGLFRCLSTSNIVLL

CpN_0064 78340 78576
No robust homolog present in Genebank/EMBL as of 11/7/98
LVMFTIQCSAQYRSRPAERAOPTQPFLARDRADFWERHPRFSACCRVLLLVAVVVAL
LFLVLMILLPLAAGSYLLAF

CpN_0065 78882 80651
CT288 hypothetical protein
YDYKYNNMFFKKNYMTDFPTFKGPKLNPIKVPNPFERNPKVARVLOITAVVGLIALL
SGVILITGTPGAPISMIIGGLCLASGALFVGGTIAITLQARNSYKAVNOKKLSEPLM
ERPELKALDYSLDLKEVMDLHSHVVKHLKLDLNLSTQREVLNQIKIDDESGLCBAA
MISENYDACLKMALYREELLKEQYQYQTFRNFQNLTHRNKVLSTLSRITDNIKAGGVF
SLKFTLSRMSRIHTTTTTLIAVSVSMVAALIPGGLILPILLVAISAGVITG
LSYLVQILSNTRKNOQDFYKDFKQVDELINQVTVTQRFLEMLKGVLEKEEVSLG
QDWYQITNAPIEKRLIEIRVYKEIDAQTKMKTDLEFLENEVRSGLSVASPSSEDP
SETPIFTQCKEPAKLRRQTSQNTISTYIGPDNENIDPEFLPMMPKKEEIDHSLPEVTKL
EPGSRLELLVEGVPNTLRELNNRIALLQQLSSVRKWRHPRGEHYGNVSYDTELDRIQ
MLEGAFYNNHLEAQEEITQSGDLVDIQRNRLGIIVEGDSDSRTEEEPQE

CpN_0066 80916 82655
No robust homolog present in Genebank/EMBL as of 11/7/98
GFMNPTQSRPPSPISIELELQELAGSSNTETISNTPPSCAATAEEVSLFIEGGR
NSEDDEGLPGSCVEYDVVCTINQDPEVRDEHVRVYINGSGRTQHEGILDAMNCDLRG
EPLFTHNSGGLSCFLGIRNRIIPRDNVISQAIQARNNEFFIAENANRDIVLFSGN
GGYLVQVADNSIYSHHILCVGIGSSYYIQGNRVHNVHVTGWTLLDRGATAVNTMT
LPSASAGGLFLPSVRCPSPYQWALQDQFNHSCHHIESPALTEYRSHQVYLPKDW
STWTRLEWIDRDSQAVLELNPQSHCRDIALYATTRISLLQECMLSVYVAPY
FMAVAVTQYSIMTLRYLLFNTNRGCRHFRFIALVALGLQSLGFLTVLLDHTNTR
VNRPPPLISVIFCTASFATGSFYVDVLTMRFFTSLSRLQLVQRRLTGRGLELRRVFN
HDSLRFSQNALITFHGGLFNPILIGFENQVLQVPRVIRPNTAVYDLNQSQAEDS
GFLAIGQITINFLCMLLVINTFFVRSVRRNLHRRPHR

CpN_0067 82920 84053
No robust homolog present in Genebank/EMBL as of 11/7/98
KSSGVSYRGPMAVEGRVNSQALNQDQCEVLANKQSKGLRLCRILSVAVITAGV
LIALTLASILTSVPYALGVFLITVLCGIFALCEKIKKVPPTPIHKEEIIAWFEER
KHEDEKEKEDPEHFGRTATDIPMSALDQFNHSCHHIESPALTEYRSHQVYLPKDW
CPVTLDPVTSSEEVILRSVGSYLLMEACVPKVSMILDELNKLKSPSERECIFZKKT
QKASFLPTOKDLATFLAYTRVNDGHLAPFRAGAKWILIHVYLRQHONQDPTFGHS
CYARLAFNQTQRLYHQLFVNEKLSIYANMDKDLPLCHPWAFPIYDILLKTEHGDGFL
QOEDREYFSRAAQDQFNG

CpN_0068 84909 84331
CT360 hypothetical protein
SFHKKFFIYSLIFSCFSAPLKGICNEDVSSQSIREDPEVLITQALVLIETPIEBGKE
IRNELQAIISDQKSSSEIEESQTSDEGLSEKTKESSNEYVLDFFDSMVQRLBGSIM
QSGSQAQIIDCFNRFDIARNRELELNRELEKLEKFLKSIDMKEKVSRELAQR
EQDIKQTLMLLKK

CpN_0069 85191 87086
No robust homolog present in Genebank/EMBL as of 11/7/98
LNFVLYVLLIFNLGIMTTPPSRSSPPPYDIELQDGNNTNNSRATPPPPEVGGELP
PYFSANFVIERGAPSLPSPQQLSLPEYSRQPPGYDETASITRTSEEMFTLVST
LCCPANSEIRDWEDHEVNCIYIASTDPTOLEAVQGGMHITELRGEPRVLYETGHLAYAF
ENTCHSRLEVSHTRAMTYFWRDFFSRHNVGRRLVFYQGNAGYVQAAALDSMHTQDI
VYGLGSPVYIRQNVHVQHYRVGRVWPSCLDSLAAACENTSVLYGESSDGIYFVPSLFH
TFDNAIRYGERCLVLCSEGGMPLPETQOQTSPLTSEGGHEVALVLPQONPEALSISR
LHHEEROGRLSNYMPGRSSNPFTMSVLRVRLNTLAQIYVMSPIYSFQSNNDIVCLIFIS
SAAVETVGYIFLTVTDSCTGRRYLRVPRVCTGLRNLALPTLLELLLSYSPRSVEGVFP
NVRFILGYMCTTRVFFAWNLILHWPRLRHGILFVHHSIIGHTIGARITDILTLASMR
YIVFVPSIVSSCLLTAHANNTLALDLYRLIESGDLRPAFNDDQMADNPWDAYSIR
GLVINTC(YMLILFANLIFMYSVRRYHRSRR

CpN_0070 87399 87208
No robust homolog present in Genebank/EMBL as of 11/7/98
YKVGFLHKNQNFSSNQSRTEYQRPKVSHPFESILQSGVGSQGTLLISFRDELKR
DLYI

CpN_0071 88066 87599
CT325 hypothetical protein
IKSLRLILEEICPLQHARCLKKQKHIIIEELFEPFGYDHLVYKLMENSSSRDAFDKKRML
KPNHVVQXQDLYLYEVQDGLLFFFTYTKALSSGIGASLFTVYSQETPSTILTCKPIF
PRLTILYLPRLNKAIEELNMRMKAIVQVLPKPT

CpN_0072 89151 89067
CT324 hypothetical protein
CT324 hypothetical protein
KPVKPKVIVYKEKVLILYCYLLFFYFFHNMSTPLDGGG(SPVQVYVQELFCDRLGSSR
ENSPUNACDSDIV:PPILVALDVLQVLPYNGHSFSTWTRKNAVEKIGFLQRMKK
YILYILAWALHAWCHITVALLTULWVGLGICV/PGIPTACTLCKENKHHVHNSLWML
UNRIIQLDPPYTRQILATNIASIALIYAVQAVGLVIGFSIQNQLSINTVYCARLGD

EATYADIRKAKKPIENI...
LKLNLQMPYCFMPEC...
IESNH

CpN_0073 89353 89574
infA-Initiation Factor IF-1
SMAKEDTLVLEQVEELLPGMHFRVLENGMPVTAHLCKMRHNRILLVGRDVTVMSS
AYDLTKARVYVHR

CpN_0074 89677 89688
tutA-Elongation Factor Tu
EDFEMSKETFORNKPINIGTIGHVDHGKTLTAAITRALSGLASFRDYSSIDNTPEE
KARGITINASHVEYETPNRHYAHVDCPGHADYVKNMTGAAQMDGAILVVSATDGAHPOT
KEHILLAROVGVYIVVFLNKMISQDAELIDLVEMLSELLEEKYKCCPIIRGSL
KALEGDANYIEKVRMLQAVDDNIPTPEREIDKPFMLPIEDVFSISGRGTVTGRIERI
VKVSDKVLVGLGETKETITVTCVEMFRELPEGRAGENVGLLLRGIGKNDVERGMVQCP
NSVKPHIKSAVYVLOKEEGRHKPFISGVYPOFFRTDVTGVTVLPCEHVMHMGDN
VELDVSLIGTVALEEGMFAIREGRTIGACTISKINA

CpN_0075 91087 91350
secE-preprotein translocase
SRSHFKQKHNRKALSRKIGTVKKQAKFAGSLDEIKKIEWSKHDLKKYIKVVLISIFG
FGAIYFVDLVRKSTICLDGITTFLFG

CpN_0076 91334 91903
nusG-Transcriptional Antitermination
QPPSCVNCMYKVVYVQVTAQEKVKKALEDFKESGMDTFOIEIILPIENMVEVKKGEH
KVVVEKIWPGYLLVVKHMLTDESILYVKTAGIVEFVGCGVPVALSEDEVSILTDIEKK
SGVVQKHQFVSGSRVKINDGVFNIGMVSEVHDKGRLSVMVIFGRETRVDDLEFQV
EEVAPQESE

CpN_0077 91956 92435
r11-L11 Ribosomal Protein
FFVSVLFEVSVQCKVRFMSVKKVIKIQLIPGGKANPAPPIGPALGAAGVINGMCKE
EFNAATQDKPDDLFPVITVYADKTTFTITQPPVSSILKTLNLESKSKIPIRNVKGL
TQAQVEALAEQKMDMDIVLLESARKMVEGTARSMDIVE

CpN_0078 92453 93160
r11-L1 Ribosomal Protein
SCRITKHGKRIRIGLKNYDFSKSYSLREAIDILKQCPVRFQTVDSIKLIDPKKSD
QOIRGAVLNGTGTLRILVFAAGNKKVAEAGADFMGSDDLVEIKISGWLFEVDVA
TPDMREVGKGLKVLGPRNLMPTKTGTVTVDVAKAISLKRKGIKFKADRAGVNCVGV
KLFSSESSQIKENIEALSSALIKAPPAKQGVLSVFTISSTMGPGISIDTRELMS

CpN_0079 93170 93688
r110-L10 Ribosomal Protein
RGMKQKTELLQVEEDKISAAQGFILLRYLFTAAYSREFRNSLSGVSAEFVLEKRI
FKAIEAGLEVDCSDTGHGVVSCGDPVSAQKQVLDNFQKHQKSLVLAGRMDNASLS
GAIEVAVKPLSKELRQVGLFAAPMSQVGVIMNSVLSGVISCVQKAGKN

CpN_0080 93720 94121
r17-L7/L12 Ribosomal Protein
VRVTKVTTSETLETVEKLSNLTVELSQLKLEEKWDVTASAPVVAAGGGGEAPVAA
EPTFAVTLDEVPADKKIGVLKVVREVTGLALKEAKEENTGLPKTVKERTSKSDAETVK
KLQAGAKASFGL

CpN_0081 94219 98016
rpoB-RNA Polymerase Beta
FREILSHONSRRTRMLKCPERSVYKKEDIDPLNLEIQIKSVQKFLQIGLAEERENI
GLEEVSHNIFPIKSYNEATVLEYSLNVLGPKVSPCECIRGITYSVTLKVRFLRTDET
IKEEVYMGITPLATDKGTFIINGAERVVSVQVHSPGINFQEKHSKNILFSFRIIPY
RGSWLEAIFDINDLIYIIRDKRRKRLATTFIRALGYSSDADIIEEFFITGESSLRSE
KDFALLVGRILADNIIDEASSLVYKAGEKSLTAMLRMLDAGIASVKAJAVDANHPHII
KMLAKDPDTSYEAALKDFYRRLRGPATLANARSTIMRFFDPKRYNLGRVGRYKLNK
IGLSDDEALSOVTLRKEDVIGALKYLRILMGDEKACVDDIHLANRRVRSVGLIQNQ
CRSLGARMKIVRERNMLDFSSDTLTPGVKVSAGKLASVLKDFGRSLSQSFMDQNPV
AELTKRRLSALGPGGLNRRAGFVRVDVHSHVGRICPIETPEGPNIGLITSLSSFAKI
NEFGITETPYRIVRQDITVDEIEMTADVEEVCVIAQASASLDEYNMFTPEVCVRYAGE
AFADTSTVTHMDVSPKQVSVITGLIPFLFELDNALRMGSNMQRQVAPLLKTEAPVVG
TGLECAAKDSGAIVVASEDGVDFVGYKVVAAKHNPITKRTYHLKFLRSNSGTCIN
QOPLCAVGDVITKGVADIAGPATDRGELALGKLVFPMFVGYNFEDAIISSEKIRE
AYTISIEEFELTARDTKLKEEITRDPNVSEVLANLGEDGIRIGAENVKPDILVCK
ITPKSETLEAPERELLRAIFGEKAADVDASTLVPPGTEGVMDVYVSRKDLKSDDE
LVEEAVHKLQKQVQVATLKTEYREKLGALLNLEKAPAAIIRHRTAEIVHGECLFD
QETIERIEQEDVLDLMPNCEMYEVLKGLLSDYETAQLRLEINVKTEVEHIREGADLDP
GRIVQVYVYASKRKLQVQDKMAGRHGNKGVSKIVPEADMPYLSNGETVQMLLGV
SMNLQOVLTEHLGYAAKTAGIYVTPVFEQPEQIRWDMHIEQGLPEDGKFLYDGKGT
ERFDNKVYICYIYMLKLSHLIADKIHARSITGVSLVTPQPLGGKAGQMGORFGEVWAL
EAYGVAHMLQELITVYKSDVSORTRIYESIVHGENLLRSQTPESFNVILIKEMQGLDGV
PMVDA

CpN_0082 97992 102221
rpoC-RNA Polymerase Beta
CSSYGRRLKNDVLEKIMFGENSROIGVLSKEGLFKLEIGIASDITIRKWSGCEIKKP
ETINVTFKPEKGGLFCEKIFGPTKDWECCKGKIKIKHGVICDRGCVETLSKVRRE
MAHIELAVPIVHIFWFKTTPSRIGNVLQMTAGDLERVYIEEYVVIDPKTDLTLLKQLLN
DAQYREVVEKMGKDAFAKMGCEATYDLLKSEDLQSLKDLKERLKTQSOARMKLAKR
LKIIEGFSVSSNHFEMVVKNTPVVPLDRLVPLDGGFRATSDLDLYRVRNRLK
AILRLKTPVIVRNEKRLQEAVALFDNGRHCHPMVACANRPLKLSLEMLKKNRNFRO
NLLKRPVDSVGRSVIIVGPELKFNGCGLPKEMALHVPITIKRLKDGVSYVIRSAKMM
IORGAEVHVDVLEEIKHPIVILNAPATLHPLIQAFEPVLEKKAIRHFLVCAAFNAD
FDIDQMAVHVPVLSVQAEQLEAKVLMADNIFLPGGKPVAPISPKMTALVLYLMADPTF
PEHGHKTKIFKDETEVLRALNNYFIDOVFVLPKDETCRGLIHIEKIKVRIDQIETT
PGRVLRVIRVPELSEFQNSMTPKIEELILVYKKGVEATVRLDILKDLQIQA
ALSMGLKDVRIEDIKHILKDAYVILVYKKGZDDIITEDERIKTISITFVSEGLSD
ALYVEIKOTRSKHNPLFMDIYAPANKQPLGALRMLAKMKNIAIETPTNPRE
GLTVLEYSISCHARKGLADTALKTADSYILTRVLVQGDVINSVQAEALDAGIETIKRS
QXSEELLPLKDRIVGRTVAEDVYQIDKRLAAGQGDVINSVQAEALDAGIETIKRS
TUNTEPRVCAKCYGLNANRLTGMHAYVITAAQOIGENITQIMRTHLHIAAITS
STHETITNSGIDVLMRLVIVLGEINILVHPLHVVGDHRTLNKXKLLKTEKIE
SLRVPVLELVKILVADGTPVQVQVQVVEVLEHPIIDCKPKIKYEDLVGIEKPKV

HLWMEV13L0SP

CPn 9083 102296 103312

L L I N D I A A V M I R Q P Q I K L I T I V C I D R F E I A A K L A D A T T I T I L L E V A G E K F
 Q E L L E A V M V G I R Q N G D L T S F I L D K I Q V N F A L I I K N I P G I S E I D L A R S F N V E A M
 V O R A V F L S Q L F E A M G K D D K R L L I P G T W E G I R A V E F L A G K I A C N T V I L N L V O A I A A A
 K A K A T L I S P F V G R I Y D W I I A A Y G D E G Y S I D A D P G V A S V N I Y A Y K K F G I P T Q I M A A S F R
 T E Q V A L A A G C D L L I T S P K L D E L K S Q H P V K K L D P A E A K K L V Q P I E L T E S F F R L A N
 E D A M A T K L A G R I F A G D I T L A T E I F I K O I A A B G A

CPN_0084 103356 103751

Predicted: Tetraodon
 LCKRIHMDYKSQLVFCPCCKGNVCFVFNLDVILTCNVCSSTYTFDSVIRNEIRQFVA
 SEMKRIHDANSILGNATVSVSVEDNQMDIPFQLLSRFVVLNLSLDGKKIAIRFLFDALN
 TSILHOESDLIS

CPn 0085 104512 103766

U1511 hypothetical protein
 F5MKFFILFLILVAQFPFASQAQPRTOVSASHSKQAKARRTSRIRSSAATNASVSRYKTRA
 AARKKIGKGFEEKKPSLSPVQWVRYSGKNYSIQTPSLWQCIDDKTQLPEKLDVLLIGKGKGN
 LTPTINIAQETSKSSKEYIEELAYHKANEMTESGIFTQIQSPSGEFTIITEKNSSW
 GFVFLCQATTVIDHTAYIFTSTATLDDYAELSFTFLKVVSSFQIRGGKEATSGDAILEKA
 LEALVENK

CPn 0086 104898 105527

ALP-ADP Synthase Subunit E
 NITMANLNADPGKGLQICDRLDLTLKPAEDEAAALLHNAKEAQAKRIIQEAQEERARKILETA
 EERAHQKIKQGEVALSQAGKRALEALKQAVENKIFRESLVLEHVTDTPEVSTKLIAL
 QQALEAQQSVSGNLTAIYIGKHVSAPRVNELLGKAVTTKLRKKSVMVQSVFGVGVQLKVEEKN
 WVLDSSSALLIFTRYLQKDFRMEIFQGS

CPn_0087 105540 106376

LT309 hypothetical protein
 LKHEKFSIFKVVVMVTQYYFLSSFLPTQLPESVPLFSISDLDLLLYLNLSENDLCNYGLLK
 LKFPDFENFAFFWAGKPIPFSGEVTQENVERMLSQQWSDDDNDFEDFFKDFLNMHKSSQD
 LKLNHFSDLFREPLSYHQTNSKFLQDYFRFQQQLRVVLGAFRRARVLNMDVSYVLRDESS
 LKPVLEVLMLQKDSPLYELPEEFSDQLQGVLDLPHLTNLRALALYQFHKLEGFCSDSYF
 LKLVILCARCATYFAIRNLSLASEVKGREINDHGLIKEAIKW

0088	106352	108145
------	--------	--------

[illegible]

Pn 0089 108111 109466

[illegible]

Pn 0090 109439 110080

DP-ATP Synthase Subunit D
 AKMSVQVQLTKNSFRLEKQKLARLQTYLPTLKLKALLQAEVQNAVKDAAECDKDYV
 YERIFYAFELFSLPCLTDCVEKSEFIQSIDNDFENAGVEVPVIREVTLFPASYSLLG
 IWLDTLSASKELVVKVMAEVSERKLKLEELFVSVIRNVLFEKKLIPETTKILKK
 VLSDRSITDVGQVMAKKKIFRLKARGDGV

pn 0031 110074 113053

pl-ATP synthase Subunit I
 LKLNHKYLFITGRNKADFFSASRELGVVEFISKCKFLTTEGHRFVECLKVFDHLEAEYS
 GALEFVGVDESVSDVIDVNLTKNEIKGLLETFYALRKLIRVRKPLGVGFSSSEIAELS
 TGLGRFFVTHKDVNDELEDSPNVYFLTAYNFDYKLVGVLDLPRDRTYEIAPRS
 VQDLVLQANLQELIRNRSNDRCLDYAYRVEYLRGLCNYDNELRQHQAECKGDELDGKV
 VAGMVTVDRIKELQSLSCRLQYIYMERVPDPTETIPYTLNCGVMGGEDLVQYDTP
 SDDKPTWVFFAFVLFSPKINVDAGGLGLFLMSLLFSWKFRMRKMKSKHLSRLMKTK
 LGGICVCTVFTLTSFFSGMSKSTFRFSTYMLLALKAKLEYQLMRPKAYKELTNEY
 LKALIRPKFALVLTALIGSAGIESAYVYDKFDINIEMELALIGVHLSGLMRKPL
 STGWLPMFMSAYLYPVLYGLTSLHYLHFVYEGGQIGYGMGSLGLAVYLAMI
 YWRGVETISVTVQFSDVSLRYETALYGLGAPPLMGRKFTNQMGARPLMLGISTITLGH
 NPLTSMQYTHGLRLNLEIYHYSFGDGGRLPPLRKLVCSGDAEASGLHNNYSIV

112121 112573

LKGAHEVEMIDMEVVVGFALVGLGLAMIGSAIGQIMAGVASHAVMSRIDEYHIGKLIQMGA
 SDQSLVYGFILMLMQAAIKNQTLSPVQGLAIGLVGAALLVNSVMMQKQCVWGLQAYA
 SGLVYKCYAAIGIVNSELFAVVEALLL

112440 113015

AAVVSAEKKMLDLRYMCSVMORLCNLHFCLLLFLRYYYKLVFGLTVLLAAISV
 LKQCNLPGLNFTETVYQPEYSAQAQLSEQSDHEVYQGVVVTWLPCLRMRKCLPVT
 NVVYVNGKVEKLTVEVNSAGYRVYCLKLEYFELQNIISYRVALCGNOEIVSRRI

CPN 0094 113194 115974

[illegible]

CPN 0095 / 115956 118790

ADNDV571 Protease K1Rinase
ACVCLDREDQDSLERYDITVIRIGKGGMGEVYLADYDPCSRKVALKKIREDLAENPLKR
LFGREAIADLHPGVVPVTYISEKDFVPLMPTIEGYTLTKLVVWQKESLSKELA
EKTSVRAIFLTHFKIKITIEVHSGRILHRLDKPDNIILGLTFSEALVDLWGAAGACEE
DPLLDIVSYEIVSSSRMTIGRIVGTPDYMAPELRHGHASPHDITDIALGVLYQMULTS
PYRRKKKGKIVLDQGRIPSQEVPAYPIEPFLSAVVMRLVADPOERYASVLTKEKDI
EHSKLGKNGWLTTLAPPKKSSSWKLNPEILLKSYPMLEVSASVSLAISNIESFSEM
RLEYTLKKNLNGEGFILLPTSENALQDGFYQGVFWLHKERTLSPVSHSLYKIEIQRS
QDLESKQITFLALEOHNNLSLFDVDTWTLLHMNVLPSRSGRGAVIVRMDIEDLIGI
FESSLSRVCLVAPDPAFLAEKLYRDLVTLVYRRIAESFGPRGKEGYEARFRAGITVLEKAS
TDNMEQEFALATIEEFSKLDHGVAAPLEYLGKALVYRQIQEYNIEIKSLLALKRYSQHPE
IFPLKLDVYVRLRSEFYSKRDRLKALVLMLEIATPAQITQGEKILVWLKDSKRAFLC
LIDPTLVYLRSSKMEFLSYWSGFIPLHLSLPHRAQDITVFNEDVLEITFYFVADKSLWFL
YSCDITFKESLEQKATIEEIVFSPFDGLAFQISTFNKDAEKIVFSDNQLSPILLV
YIFDLNAPLSESGEAIQALDILRSKVPENFYHDYIPKNEIEIRAHFWKNEKALSTIF
ENYETQKLDQEQHELPLYGVCYLALIGQAEAAQKHQFVCDREDIRFPASLLARNYNRLGPL
KDALSYQERRLLKRLFYFLCYLGNHDERDCTMYHLLTEEPFL

CPn_0096 124347 118837

CT296 hypomethylated protein

CT296LIREFPMKSLPVVSGIKVRNLQNVSIHFNSSEIVLTLGVSGSGGSKSIADFDTLYA
AGRRKYRIITLPTFTFATTITLPPNKVEYIIGLSPSTAIQKMFHSHYSTATGVSTTFELFS
AAALIFTLGGQARDPQKTEVLDFYSKEVLSITMELSEGVSIILAPLRLKDIAIATHEYAQ
GVSLFVRNCNCTIHFTSQFSGIPEDCSVDFIADLRSKNRIARKGVLFTLAEFLFEGDH
GCVLSEDEELTIFSPYQKIDDDVTYPLTQQLFSPHAIERSCNLCQSGKSIITSIDNPLLH
LISIKENCSQCFAGNCSYVYRHTIYQALADALNFNLTPWKDLSPEIQNIFLRGKNLVLV
RLFDQTLQGNLNTKYYVWGLINDGKVRYSITKPSRYLSACSAHSCSLCKQGLGDTYA
VATWEGTKTTFEQQMNLNNHVVFSFKVSPSLSIQELIQLGKQRLSFLIDLGLGYLTPN
ALATLSSGGQERTAIKHLIGLHFGKFGITYIIDEPSILGHLPDTEKLGIIVKLKRDQNTY
LVEHEMRISLADRIIDIGGAGIIFGCVLFGNPKPDFLWNSSLTAKYLRQELTIPIPI
SREAPTSWLLLTTEATHNLNKLITPLRLARLIGTVGSGSGSKSIINNTLVPALRESFLK
ENPNKILHFEGCIGRLTHITRDLPGRSQSIPLTYIAGFDDIRELPASQFNRLOGLTK
ENPNQPGQACIQCGQLGNTISDDTPIPCSECGQKRYHSEVLEIYEGKNATIDLM
YAEAEKFFISHPKIHEKIALCSRLDYLPLGRPLSTLSGGEITQRKLKLAHELLFASPKQ
LVYLDPELTGLHTHDQALTEVLLSTLYGHTLVITENHMVYKCVDDILFEGEGGDL
GVYGLASCTPKDLIGLHTPTAKRALPYIESGLDTPVVKSEPPSPKSDYILKDAYONNL
HITDALPNSLITAGPGASGKHSVLFDIYASGNITAYALFFPYIRQGLLKECTPLPSP
VEKGLSPVTSVIRKCSNNSRYSITASGLCSNLEKFLAYIEGFPSPTEEKLSTKSI
IIDSLLSKYKDDYVITSTSPILGSDLEIFLQEKQKQFIIKLYSEGNLYDOKRPLNLII
PAVIAQHTKSPKNSSSLLASVAFSLSPSSEITWIIYSQKORKLYSLGWKDDKRLYP
ITHOLSSDQLDEPGRCRLTCGGREGLKISLEHKEKIAHYHTFLEFFSFPKSPYQVK
LKDNASQPLKLTTFKFLNFCRGSSEFGMNALEMLQDTEESDPLIKPMLLTCSPA
KSGTGLNDYANVYRNNTSLLDIYQEDATFLFESLNTIGTDDTRSIIQDLMNRLTIFSKV
LSYLTIQRQDPTISDGENYHLHAKKISINLTNIVYLFEEPSGLVHPQDLPTIQLLKE
TIVANRNNTVIATDRSCSLIPHADHAIILFGSGSGQGGFLPEEDSLDTEVCPSDVLANVPOTEV
PKAPKLSIKCANHTGRSDRLKVLNLSHHIQNLVKSAPHALVAIGGVSGSGKTSKALNI
KKQAEILLAKGTTITSDVLTIDNSHIFASSQSDISTYFDIAPSLAFRAISYQAKALNT
STMFSTNTKQGCSDCGGLGYQNDIRAFYALEKRCPCTSGFRIFQPLAEVLYYEGKFGH
LLHTPTLFLIDELFSPKIDIKQHLPEKLRSINSQHSVIYDHDVKKLKSADYIETGP
SGKQGGKLLFGSGSPDIPYASKDLPEKLYICNEELDS

Pn_0097 124549 126006

SMG1TRTKIKITIGPATNPSPEMLAKLLDAGMNVARINFSGSHSETHCQAIGFLKELREQK
 VPLAAILMDTKGPEIRLNGIPELPSVQOQKRLRVSSDIDGSAEGCVSLPYKGIFFPVE
 IADLDIDGGYTHAWVSSSEADSELEPMENSGLLKHKSLSLRGVDVLPTEKTDADLK
 IAEONGDVAAASVFRYGEDITETMRKCKADVGKCPMTAKIENLRGVENFSKIAKLADG
 IARGDLGIELSVSEVNLQKMKARVSGRTGHFVVTATQMESLRNVLPRAEIVSDY
 IASGDSAVMLSGCTASCAHPVAAVYIMRSLVLTENKSHDSFSLKLDLNSGALQVSPY
 IAYGLAGIQIAERAADKALIVYEGSSPGLSEYKPKFPIATVPSTSVYRLAELG
 PMLTGESDRAVVRHQACIVYIEQGLINZYDRILVLSRGACMEETNNLTLTIVNDILTG
 CFPTET

pn 0098 127494 126091

[illegible]

127527 127804

1AVPAKSPADIVATFALDEELSEF007TVLJAAKJ00PQJ0KH1KFF1TRF

QKYMQI LLDAPVSLLYGAF [REDACTED] INI NITSENI ANSTKAAEEEA KRYVEEKGRGFETY
WEEAKORLEATAAELDDLRL [REDACTED] EQEIPLANLKI SIFSDLNLRKVSVEKAAEEEIQ
GIQEYAEHQI EDLELQKQFEDLQKKLEALERLLQIGRRIDSSVDKQKELLGLLGREE
AA

CPn_0151 194178 197625

[illegible]

CPn_0152 195274 194318

LTI49 hypometicLSCF
 LLMKRVKVAISLSCFVSAIGASAAPVRVPGFPQIPEDLVQIKTEVCQKQEVCLATIKCD
 DHNLGLVLPHLPTPTBGGGFTLVLEPHFGRTGLKRGKFAAVGIALTRVMDL
 AGCGDSGEAEVPEIPTYLRDAQTIVLETOEHPLDANALRGISGFSLOCHIAFELAKIYN
 PRDLNIKALSWAPIADGGGLHKELLYENFKHGGBDIVSVGKDFGGPGPPPIVCSGSDVLE
 LIRIQDHWANSLPTKPYLLHQOIGDDTLVSRTOCTLFKNTAPGRMTFISYPNTGNGSLAT
 APDLMDLQIVSHVFORTL

CPN_0153 / 195430 197892

teus=Leucy" tcrna splyncase
NMRYPNLDIEKKWQVKEHRSYQNEEDKVKYVLDNFPYPSGAGHLGHGLIGTSD
IVARYKRAGRFSVLHPHMGDSFLGPAEQYAIRQTHPKVTKQNIANPFKKLQAMGFSYD
BGRFATSDDPDYHYTKQLFLVYDQGLAMADVAHYNPELGTVLSNEEVENGFSIEGG
YVPERKMLQWILSTIAYADKLGLDLDALDQPNVQKLOKWKIGSEALDTHFLTQEGS
LEAFTHLDTLDTLGSFLVITAPEHDPDLSISVEEQREDTAYVOESLRKSGRDNISVTK
TGVTGNYAKHPITGNLLPWISDLYVYLGCGTGVMGVGPRAEDTFAEMFSLPIHEV
DNGVCHSHTYVTCNLSGQAEQDYVNIENLRSGLAKTHYRLRDLWFLSRQRYGEP
DIIHFDEGTHAPLEDELPAHLPNTDIDYRPEGFGOGPLAKAQDWHYIYDEKTRGPRCE
TYTNPWAGSVOHYLRFCDNAHNSLQSKERESYMPVLDYIGAEAHYLLHLSYRFFWR
YVDAGLVLTPEPFKKLINOGLVASSYTIPOKGYVISIEDVRENGTWTSTGCEVVERQ
EKSSKSLNGLVDPQVLEIYEGADLARLMEYSGPLDNKNTSNEGVGGRNIRNFYDL
TSEVQDIDRGLVLAHLKFLITHTIEMKSLNTPSSFMEFLDNFSLKLPVYSKRLSM
AVRLEPIAPHISELWYILGNPGPQIDQAAPQIDESYLAQVTVYVQNGKLRGLRE
ACAPKEVLSLRSVYAKYLENAQIRKTIYVPNKLNVFL

CPN_0184 197874 199202

g95a-b100 TRANSFERase
 TSEDFPMLRGLQVHRILPFLVLDVLCVAFILPKLLYKMLVQYKKYKSLAVRFLGKLKPHV
 PGDGLDGLVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRVLRV
 LRLVDFSLIIKSVAWLKRLPSLVVSGDFQWLNFEAAKRGITLTVINGRISIDSRSKFR
 CILRLQGLQVSPVDGLQDEVQKORFLSLIGLPEHKHQLQGNITKTVLYAAQTALHLERET
 WRRLRLPTDSDLVLVSGMHRSDAGLGLPVQKGLKEGVSLVWVRPHRIGTLDVSESLR
 LHLVPGVLSRGANFSVPMVSDVIGKLLQLYVAQDLFVGGTDFDPKIGHNNLEPQCE
 VTLFGPHITSQSELAQRLLSSAGGLCDETEPIIDTVSFLNANQEVREAYVQKGKVFV
 DETASFDRITWRALSKSYILPYKNS

CPN_0155 199697 199488

NSISFGVPFLEKLKISLPIEEMRHLEFMKTHNSSSNGFSNQEKGIRTYFKSDLLGYEDL
YFLRENINPN

CPn_0156 200147 199770

NO robust homology present in Genesbank/EMBL as of 11/17/97
LGKQKLLARHHMHNIVLSEEPGRSAFLGRATFFPNKYPIACGGVGIPSTIGNLFTIWC
FYFYRAATPQSDHPDGGCFILLERLKLKGAGFFYCDLRESNTTGFTLFFEGSNKGVLNKH
LFI RDE

CPn_0157 . 200753 200298

No robust homolog present in GeneBank/EMBL as of 11/17/98

CPn_0158 201463 200894

No robust homolog present in Genbank/EMBL as of 11/7/98
PPKITSINIDLLLEDLDTDSIPWPKLYLSEDFDFAYFPESKAIIDTVAKLEKNNPGEF
CLESKKILARYLLEQLFKLETGLNFPSTIDGGRESFLIEFSHETKKKPTVWAFIFYFYHH
SNGPKLEKDFQACCEVHNRLNLGLK/RPQAGA/QNDGRNGGPGYIGFLIWEENYGSV
LKDHGFIKDN

CPn_0159 201811 201467

No robust homolog present in Genebank/EMBL as of 11/7/98
CCFQGETATRIFSMTPSGFSLATEEK/QVSTAQKVIKILALIFFPIILIALAIRYFLHRK
FDRKCEVIPQDTPKELELILANPOLVEKAAREVHPGFFALPTKYQSMYIQTSKG

CPn_0160 203794 202127

pfkA-Fructose-6-P Phosphotransferase
TVLLSLNLSYFEIQRVLRVLEITLTLTKSHIKIQTSSPPSPPELQKHINPLCRIP
VLYIYTEOSTSKPLIGVLLSGGQAPVHNVVIGLFDALRVNPLDGLGFIKPLSLJTR
CLYKQLDLSVYDDYNNQKSLSSPEKIEITEQRKNILNTVKQLKDLGILLIGNNSN
TDTAMLAEYFLRMNGQSYIGVPTKTDLNNKWETELSGFHTSCTRSEGMINLAKDL
KSGKHIFHFLRMNGQSYITTECGGLTLENLIALISELIATRKLSQLSGLQALGLVRR
KSGKHVTVLIPGLTEHIFEDTRKILDELNVLLANDGSDTEKILSKLSPELTKTFHLPK
DANQLLARDLNRVNRVSKLATEELLVAWVKKETEKTKPMHEHIFVYHFFIYEARAGFP
SNFTNYGIALGILNALFLVRQKTYTTLTNNLAKSYTEWQGGATPIYKMHLENRCETE
TSPVITDSVPKSPAVQHLLQGGDGLLEDLYKFTGIPQYVKEELIDRPLTLIWNQET
HSFVQALYTPSKGRLL

CPU_0161	201658	201798
----------	--------	--------

(predicted acyltransferase family)

TMQKGAITYAGMTLNPDFYQFLKIDIVKELMPSAR... YMOGEQDLLVSNHRTL
TEAFANQDKPITILTYPDVDAHFAEASSALSDLT... TSGE

CpN_0162 205870 204803
No robust homolog present in Genebank/EMBL as of 11/7/98
FVYLYNIQSPFRIMKLYSISDDVDTPWIFQMLSKVDSYFLQGNRIKVVIVMOEPNLI
IKGVENVRIGTIVKILKLSLIFPLILIALALHYFLHAKYANHLVSKILERAQVYPI
NVRSGDTASHYKLTTLVPVQKNLQAMGSPLEVEAALRTTKPSFFCVPAKYQIIISSH
...
PLDEDRGGGFEILEQLQELGVRFPCIPQSGPDNPFGQFQIRIYWEDSYQPNKEV

CpN_0163 205831 206394
No robust homolog present in Genebank/EMBL as of 11/7/98
FEKAVYCIKCKQIIKCSIIHTPTATPLCTEGEIFPGVDSAIQNDLERLLTVKKRPD
IIREYLRAGGSLVTTYKQGRRLSPQLRVLDLQVSPNHLAELDCAIPQDLIGA
TYIITPADFTSYILSLRSYQANSPPDDTWIGFWSIDDPQVAVISFLKHGFPALSTLAQ
DPLCTNK

CpN_0164 206444 206998
No robust homolog present in Genebank/EMBL as of 11/7/98
LCFKCIYIKIIFSLKQMLTRSTIESSDLSRSFSQKLSVQTLKNCESRLMKITSLVI
APLTLIVGGALIALAGGGLVSLPLGLILGSLVLFSSYILVSCCKPFTLKEMTMCVSKS
KINIWFQQRNKDIEKALENPDLFGENKRVGNRSARNOLEMLIHETDGLILKRYMKGAK
MYFYL

CpN_0165 206983 207582
No robust homolog present in Genebank/EMBL as of 11/7/98
NVLLEPMWPKTIDHVPESIEDIRKVVSCYKLIKEQEPFRSLISELGLVIRCGRLRLK
RSKYQEQARTVSDADPLFCLTRSYQDGLTPLRAGPRDLINHYIHLRRRNPKHFFSP
KHPCYARLAFNESVYVRELFDIERLTKMYVEGDYSKEQEKQLAILSVKTLDEGKDF
LIEHKDLDLGRGFTDVCT

CpN_0166 207594 207962
No robust homolog present in Genebank/EMBL as of 11/7/98
NCLQKYNKSDSISMSINRSIHLEASTPFFIKLNLCSERLVKITSLSVLSLLALVAGVT
LVVLPVAGILPLLVLILEILITVLVLLFCLVLEPILIEKPSIKELPKVDLSVYED
STL

CpN_0167 208309 207977
No robust homolog present in Genebank/EMBL as of 11/7/98
NLWGHFPRGFMLPFCPTILLKAPFLNSENYGLERLAATVDSYFDLQSGQIVFLSKQDQ
ITVEELSAKDRKFKPGSMNCTLYTEDPILPAHNSFNCSDIQMRTPIPIH

CpN_0168 208716 208417
No robust homolog present in Genebank/EMBL as of 11/7/98
SYLRLRRRENPEHFFNPGHFCYARLAFNESVRIYKLFNTAELKQMYGAGDYEQNEDN
LHSLISFVQILDEKQFDDFLATHKDTTIFIRGGADIFCS

CpN_0169 209537 208710
No robust homolog present in Genebank/EMBL as of 11/7/98
SRHFEPTIGENMKNGVSGSCSPVMEANTQPLRNLCSERLVKITSFVIALALVQGLIL
TALAGAGILSFLPWLVLGLVLLVLCALFLFSYKFCPIKELGVVNTDSIQHNFQKRN
KDEKATENPELFGENRAEDNRSARQVKTETLRDCGNVLLKIIYERNLDVLLPNSPK
TMDVDPVSEDSIRTVISCYKLIKACKEPFRSLISELRLAMSGGLLRSRYSYCAKT
VSHKADPLFCPTHSYRYYRGLTPLRAGPRYINRAI

CpN_0170 211098 210025
No robust homolog present in Genebank/EMBL as of 11/7/98
NVRKNHIIIRGEKYNTCTVIAFVLSYDITLFLNLEKEDSVHKICNEIFALVRLNTIAAT
EALIKNLKPADIHVHLPGTITPQLAWILGVNGFLKWSYNSWNRHLLSPKPHKQYSNI
FRNFODICHEKDPDLSVLQNTILNDFNSFDRVMATVOGHRFPGGIIONEDLLIFNNY
LQKELDDTIVYTEVQONIRLAHVLYPSLPEKHARMKFYQILYRASQTSFCHGITLRLFLNC
FMKIFAPQINTQPEAQEAQVQLQEVSTYFGLFVGIQSGASEAPGACPKRLASGRYNA
DSGCEAHAGEGIEIETRTIFSSAKVNEGLIEITRVTFSSLRKQPSLPIRVTCQLG

CpN_0171 212444 211149
*guaA-GMP Synthase
IKLQASRRHLNTIFILDGSOYTYVLAKQVRKLVFYVEVLPMNVSQCLKERAPGLIIL
SGGHSVYENKAPHLDPEYIKGLIPILAICYGMQLMARDFGCTSPGVGEQYTPHLPY
CELFLKIVDCESLDEIRMSHRDHVTTIPEGFNVIASSTQCSGIGENTKRLYGLQFHP
EVSOSTPTGNKILETFVQIECSAPTWNPLYIQDVLVSKIQTVIEVFDEVAQSLDQWL
AQGTIYSDVIESSRSGHASEVIEKSHNVGGLPKNLKLVPLRYLFPKDEVRILGEALGL
SSYLLDRHPPGPGGLTIRVIGELPEYLAIRRAADLIFIELRKAKLYDKISQAFALPL
IKSVYKGCDSRYGTYIALRAVESTDMFTGRWAYLPCDVLSSCSSRIINEIPEVSRYVYD
ISDKPATIEWE

CpN_0172 213237 212440
*impD-Inosine 5'-monophosphatase dehydrogenase (C00H-terminal
region only)
APIGAAIGIGPLGISRAHHLVEAGANVLVIDTAHASKGVQTVLEIKSQFPQISLVVGN
LVTAEEAVSLAEIGVDAVKVIGPISCTTRIVSGVGPQITAITNVAKALNSAVTVIA
DGRIRYSGDVVAKAAGACVMLGSLLAGTDEAGDIVSIDEKLPKRYRGMGSLGAMKOG
SADRYFCTQCKKLVPQGVGLVAYKGSVHVDLQILOGIRSGMGVGAETLKDLTKAS
FADRTESGRASHHNIYKVQPTLNY

CpN_0173 214041 213745
No robust homolog present in Genebank/EMBL as of 11/7/98
TIFOLYKIDSYKHQCFMDSVFPDRFVSTSPSIDIAKTLVSNCCHYCSRCLFIF
LGLLSIIICFNVYTSGETASLVFGLISLIVLVLLIIECRNRECCRRIS

CpN_0174 214215 214724
No robust homolog present in Genebank/EMBL as of 11/7/98
PIIPWPEKIVLSMIMTTPISNSPVALNELSLIPPTLVGDTTQSLATIPAGRRS
TLRIILDIPIIILGLATIICTPIVIFPLANGLLSTPISIGCCCLIVGLFLMGLYFM
IGSLDYLAVGLLQVLSQAGREDEYIQEIALRGAPRAECTPESPSTWL

CpN_0175 214446 215275
No robust homolog present in Genebank/EMBL as of 11/7/98
LILACGPIILKPKQMGQINCVIQDTTIVYALNSFDRRLSDGTHIRLOKQSPLEANEALGE
FIEGLIATEPFLIEVATPILQYIPKLYLSFIDRGGVHYEVLDGVFLKVAACIENS

FLTDGMSPELLSEVKEALM

CpN_0176 215271 216518
CT153 hypothetical protein
NDDPMDSEDCGEEASKDSAFSASFSYEFVKSTRESKNTVTHJASRTLYILRQDCSYDP
RALVKVDEDFRYWVKRLDAKNPDSINAFVKEVGTHTYASVYGGIGFQVLMKSYLQVEEL
EKEKISISVAAASSLLSKSTONATEKCYSSYQSESSACTVFLGGTVLPDLQODKLDKDW
SEGINPEIPALASVSSITDLIPELFSDEDAQVLSQKSAQVILNLYESHKPKKEEGP
...
LEKLATQDGLFIRSGDEILKHNTJOKTANTJMLJQKLTCTQTSUJVFITV

CpN_0177 217513 216608
No robust homolog present in Genebank/EMBL as of 11/7/98
DKREQTKSKFIFLISESMKQPMSLIFSSVCLGLGLGLSSCNQKPSWNYHNTSTSEFF
VHGKNSVQLPHYPSAFRTTQIFSEHNDPVYVAKTDEESRKIWEIRHKNLIKGSYIPI
STYGLSMHPKSAALTLKTYRPHPIWINGYERSFINDTGKYLKNGSRRTSHDQPKNRAVL
NLKSSGRRGNAIGLENTEDFVIARRREGVSLVPEVCSYPOGNPFVAYAWTADESA
CSKEVLVPGYSLVWESVSSDSNLAFGDSFAEDYLRSTFLANGTSILCVHESYKVPVP
QP

CpN_0178 218052 217789
No robust homolog present in Genebank/EMBL as of 11/7/98
VKEYLDLFLVQRNVERDPQTKRHCTVSKQFGESIDAKTTTQGLFIACKTEPHGKCLGL
ESILKQLALGITTGYENREVEVWYLD

CpN_0179 218550 218056
No robust homolog present in Genebank/EMBL as of 11/7/98
PKIWTHTFETRIEATSVKFNRLRKSPHKSQRRSRSPKACVAFNFTLQAGRSIIPG
KKAILLNVNDAKTPNYSCIFESIGFFNDEQAQNALVRLKLVPHHFLKGLIAK
LPBSLKKDRKFMSSLIPTKLSYALDLSAPMHLEKPNLSYEKLD

CpN_0180 218963 218355
No robust homolog present in Genebank/EMBL as of 11/7/98
LSLHLILDCYKPVFIQNTVASEYTPSQILHAQREVDAYFNQADCHPARANQILEAKKI
CLLDVYTHNYSVPTFCVDNPNRLRFTFVSSKNQNMGLSNDLVLEAMVRRHTLAK
LAACKIRNIIEVPRVGLDLRSGLISLKLKQPOQSLTDFVNHSTNOEARVHQKHL
LISLILLCKQAVLESFQEKRRS

CpN_0181 219175 218777
No robust homolog present in Genebank/EMBL as of 11/7/98
VHELKIDGVYFFKFMKLFYNNYSLSHHEKSSLEKAVQALDSYFYWGDDTDLVAR
DDISREIYCVRLYIRFWIVSISQSLSRIPWRLKRILLRYCTLRGKIVMFIKRIAILL
GLIRFSLRKSXY

CpN_0182 220704 219334
accB-Biotin Carboxylase
RCIMKVLIANRGEIAVRIIRACHDLGLSTVAVYSLADQALHVLLEADEAICIGPOAAK
SYLKISNLAACEITGADAVHPGYGLSENANFASICESGCTTFIGPSSSEIAMMGDKIA
AKSLAKKIKCPVIGSGEIIIEDESEGLKIAEKIGFIVIKAVAGGGGRGIRIVKEDEFY
RAFSAAAEAEAGFNNPNVYIEKFINPRHLEIQVIGDTHVHLGERDCTIQRRLK
IIEETSPILNAEIRVKGVAVDLARSAGYFVSQVTEFLDKDKKFFYFMDMTRIOVHT
ITEVTOGIDLVKEQIHVMGNLPMKQKNIIEFSGHIIQCRINAEDPTNFSRGLRDVY
LPPAGSIRVDGACYSYALPPYDSNIAKVIATKGNKREATAIMKRALKEFHIGGVQST
IPFQFGLNPKFLESYDINIDNLLAOGNSFFKEF

CpN_0183 221207 220695
accB-Biotin Carboxyl Carrier Protein
RRIGMDLQIEKLIAMGRNGKRAIKRGELELERDREGNRQEPVFDYSLRFGSGS
QERPIPTDKPKDTIKETTTENSESTSTSSGDFISSPVLGTGYGSPADSPSPVKPGDIV
SEDTIVICVEMKVMNEVKGMSGRVLEVLITNGDPVQFGSKLFRIAKDAS

CpN_0184 221814 221221
eIF-Elongation Factor P
QWKIKFCCEEKIMVLSQSLVGMFISTKGLYKVTSVSKVAGPKGESFIKVALQAASD
VVIERNFKATOEVEKQAFETRTLEYLYLEDSYFLDLGNYEKLFIPOEIMKDNFLFLKA
GTVSAMVYDNNVSVLEPHFLELMVSKTDFPGDSLSLGGVKKALLETGIEVMVPPFVE
IGDVIKIDTRTCEYIQRV

CpN_0185 222457 221765
rpe/araD-Ribulose-P Epimerase
AEVKQESVLCVPSIMGADLTCLGVEAKKLEQAGSDFIHDIMDGHFVNPLTFPGIIAA
INRSTDLFLVHAMINYPFEIESFVRSGADRITVHFEASEDIKELLSYIKKGVQAGLA
FSPDTSIEFLPSFLPCDVVLMVSVPGTQGSFLPNTIEKIAFAHAIKTLGLKDSCLT
IVDGIQQQSAPLCRDAGADILVTASYLFEADSLAMEDKILLRGENYGVK

CpN_0186 222878 224068
*similarity to Cps Inca
PIKDKILMSSPVNNTPSAPNIPAPTTPGIPITPKRSSFIEKVIIVAKYILFAIATSG
ALGTILGLSGALTPIGIIALLVIFVSMVLLGLILKDSISGGEERRLEEVSRFTSENOR
LTVITITLLETVKDLKAQDQTLLEIENRNEQNLKTTAEDLEEVSKLSEGLEALERI
NQLIQANAGDAQEISSELKLLSGWDSKVVEQINTSIQALKVLLQGVQVQEAQTHVKAMQ
EQIQALQAEIILGMHNSALQKSVENLLVQDQALTRVVGELLESENKLSQACSALEQIE
KLQGHETSQQRIDAMLAQEQNLAEQVTALEKMQEQAQKASEFIACVDRDTRFORRETPP
TTPFVVGDESGEEDGCTPPVSQPSVDRATGQDQ

CpN_0187 224218 225045
predicted methylase
VFLTYTRTLPMHKSFLSRKKNSHKEETQNDTASSYNIKIVQDKGHYHIRETILPOLLP
SLTLQKSSVLDICCGQFLPALPKECRYLIDISPLIALAKMNRVNGHOFKAVDLS
KLEFVEPTLFSHVAALSLQNMFFPAINTANTLLELPQOFLVNLHPRIGRASSV
HYDEIKKATSRHIDRYLSFNKIPIMAHYQKQDPTLQPHFPLTWFKELSHGFLVSGCL
EEMTQKTTQKRAKENLAKKEPDLIMKIKIK

CpN_0188 225090 226400
CT152 hypothetical protein
NRLKCEIMFRKLPFPKSKKTQQRORLNNHLLQALIQKVLIIHACKEACVLSYGLL
TVVILVFLRLDNLILNLAWEELIKFHYKKTVAIVAAVHATEENIGLVIVGCF
VPPWALIMLLSLEDLNLKIFETMPLGLPIVAGIIVLQIIMFIVGEMWYITQ
IMFYAGLFLSLHMTALYICRPVYLLYLLALPCTAPLPPVATKTALEKTLTIG
VWVYQKAFPLQVSIYNSPTGALVALPFLILLVYTMVLPRAITPIIQNRCT

HLGNTVKKILK/LGSPRSLRTTSLWYRP

CpN_0210 252459 251440
No robust homolog present in Genbank/EMBL as of 11/7/98
YQKWEREYFKTIREKEHATISTMLVEALKEFAHLKQKPTSDQETISLYQCLDH
LEFVLGLGQDKFKATETEDVLFESQKADIAWIALTKARDVLGLGIGAIYQTEFLG
AYLGVNRRAPFCIAIEIHFILKTAIRDLNAYLLDFRWPLCKIEEFVWGNDCVEIAKRL
CTFEKETKELHESLREHEAMEKCSIQDLQRLKLDIIEIHDVSLPQFSKTPSQEEYQKD
LQVLPPLALLLQETLKTNDPLAAYKSEIRKPHALRYKIPOTKELREAI
APVLEKEL/MPVYKAAHILCMFREPTTIFKDTK

CpN_0211 252765 252463
No robust homolog present in Genbank/EMBL as of 11/7/98
ECVMSYPDISNVQASSIQSALLHKTSDQIQKRCFKQSTFVILAVSLVIIGSLFLAGVA
ILTVFSGVLSVFGVLGVLGLLLLAGGVLLVEAKSL

CpN_0212 254081 252888
No robust homolog present in Genbank/EMBL as of 11/7/98
ELSYGVVSIYSEILSFSELTSCKHSFPFQPIETASIRIHHVFNVIIVLILGLTLFVC
LGMVFLGVSTYLLGSMSSILGLLLISIGLALKFKERYGLEPKELFGVGGDFDKLPSE
IIOQMDQADLARELDLEQKQDILRGFSARLDVLEGSKEKKQILKIGVPRNLSEIQER
AQONSILEQCKEALLFRKSAQEIFKKLYDRKAARFWSYREDLWCYSEIHVSCKNSL
YIGDVFEGTAPHLEAMCYAMCRATKNLRNYYVCVDMRVNEEKKRAKQLSVSELCCT
EIEDLENETNLFTSDSDVLEEQYHCIRVMTLHALWAYNDEVSRKPIDTLDVRAR
MAVEDCIEFTELQMCVHTKTELEIAQYVVDILEA

CpN_0213 254345 254190
No robust homolog present in Genbank/EMBL as of 11/7/98
ILVFSRVIFSNTNIGIPRELILPLMKKENDPPCFLSRVEGTFIILNIK

CpN_0214 255768 254446
No robust homolog present in Genbank/EMBL as of 11/7/98
FLGLKEDYERTYCNIPAPHPORVDSKGGCASHSVTVVVALFILGIFFLSGSLAFLVH
TSCGVLLGAALPILCIGLVLLAVALVLEFLCHKHTRQDLDDYDQDLVLHKKIEFMDI
SELVTFEKLQNLQFHTKQDSDLQELQKGFINCMKWLLEDEVTKFLIVDRFLFETR
RNFTTFEGQVKGQISNIFDLHEEKSLLYELYLKRDQLVLLNFFLLPGGILKVDYDEIE
ATKGLFIRLTSRLDKLDVKAQERKKFINEMSEFKEVEKAFDIVDRATKMLDRAKES
ARLPMGRTESSLEMKNEEALNQGLDPENLSHPELSPYQQLLILNANSEIVLHYEF
LISGTVTSGLTLEECENRMAASTGLNALLVRKLQFRGAISAYFEKLEIEKELSLQD
VKSLELELHKIKDIYET

CpN_0215 257039 255759
No robust homolog present in Genbank/EMBL as of 11/7/98
LTSSKQVMSAIARDCFSPSPQPSSTLGVHPKPKYSLILSVSLVGLVLLCGVMLLL
VNFSSFSVLT/VGLGGAGVFLGSLLLILGLIFFVSYHRKLSAATRSLEQKITLEVQPMAD
LRKLEINQVQWNSNLLDEMEDFKEVVAHQKSFATFEGDILLFGREVEKYETIWKELDGR
DVALLTELKINWPLEFLRKKQDRLQCEIDKLKREVMKQKSGKLKACELTKFSALKDQ
KLEQCYRDKRVEKLEVPFEGYRRELLEVLKTRLSVEKEIQLEFVSAFEKSLASLR
TVFSEELQALDQAKAELLDIQVRKSVVDELSCPTLIQYHLLRLYEVCRIVEQPLTQ
TFSQEQKVEEYALKARIRKTRVRLQVRAVAVASTDILSESELGDSNDFED
ANDDFLD

CpN_0216 257623 257174
No robust homolog present in Genbank/EMBL as of 11/7/98
NKARTNMPVTFDRIQVDFIPEDTSLRINSYIVAGGLILGVVLSILSVICLDIGLVLSA
GAFTLGLGLIFALFLFSFLLSQKVRPDVLSYLEKEVPQYETPLYKEDLESER
DMSAISERLGIIEEKLRIAEFRYSDSV

CpN_0217 257881 258579
No robust homolog present in Genbank/EMBL as of 11/7/98
PKCKLKGFLSVNELIFGQTFQSVVVLGVFFASRGKAWLTQWLSLLSSIMNVFLKQHL
WGFEVTSADVYVIGLLTCLNARYEHKNDINDAMLCWSVISAFLVLTQHLFLISPN
DSSQEHFLALFSSSTPRIVASLVTLFVQIVDIKLTFTLQVRFSSKYFAMRSTISILFSQ
LDTIIFSLFGLYGLVSNLDCVMIFAMLVKGVITLAIPTLVTKAVLDRSS

CpN_0218 259064 258582
No robust homolog present in Genbank/EMBL as of 11/7/98
IFLSKKVFFESYEDFAMVASSWPKSLRALVQGRYFVDELKETPYRIHDFKPTPIHRLY
RSLPIISTIGIIRLIEAHSGPIHPRDKMYRFEVLQAVIEILGLVLVILFDIGCFLA
FLVAITLSLLYCNSTFTCVNLSPFERMLEGIGEAIVFLA

CpN_0219 259348 260472
tgt-Queuine tRNA Ribosyl Transferase
GSSALAKFHLIHSKKSQARVGGIETSHGVIDTPAFVPAVATHGALKGVIDHSDIPLFCN
TYHLLHPGPEAVAKLGLLQHFQGRQAPIITDGGQFIFSLAYGSVAEEIKSGKKKQMS
SLVKITDEGAWFKSYRDRKGLFLSPESLVQAQKDLGADIIPLEDELFPHTDOEYFLTSC
SRTYVWEKRSLEYHRKDRPHQSMYGVIGHGLDPEQRPVIGRVFEDFPDGSATOGSLGRN
LQEMSEVVKITTSFLSKERPVLHLLGIGDLPISYAMVFGGIDSPDSYPTKAARHGLILSK
AGPIKIQQKYSQSDSTIDPSCSCLTCLSGISRAYLRLHFKVRPNAAITWASHNHHMQ
QVMEIREAILKDEI

CpN_0220 260660 261238
No robust homolog present in Genbank/EMBL as of 11/7/98
FYSFLKKGKIFYMSKESIRSYSEISTPTPIFRETSPKEGVAYKLQLRSPAKDCILNRVS
LKALLRSIPFYGSLGAKRIHSAWAKDAPCTTRV/HYVGGLELLGLGVVVLACKVLA
TALKFLSKASSIKQMKWREKARNLAADTKQSVOLTSCTFRCFLRNRVVEE
GASEQTVREIV

CpN_0221 261621 262064
No robust homolog present in Genbank/EMBL as of 11/7/98
DALRYKYIIGQVNRKYKSAEFSADHYDDNLRVYKRNLRGLAPVENEVCLFEENNL
LEVMASIPMGIILGRLHLSWGTQDFKDSKEIIFHTALGILETLGILVLLIKIT
ITILLIFTLTLLCYMYCAAYSDPHI

CpN_0222 262474 262442
*weak similarity to Bacteriophage CP1 (Oct4)
GEKFLIKWEKRLNFAELTPQPEYRNRVLMNLYKRFRTQAKVWYRCVHEAGLYE
KIKPLTILYDDKHLQVGGVLRKHLQGLFLRKNLIPKIRYFBAAYCTKQRPYHL
LLE

CpN_0223 262950 264444

No robust homolog present in Genbank/EMBL as of 11/7/98
ATMLIGRYSSDQFTEATKIKLGFVRDNLGLTNPISIVSETSSIKDSVLRSL
PIUGSILGCARLYTLTNDPLDETEQIMWITFGALETGLGILILLFKIIFVILHCF
HLVICFCK

CpN_0224 263402 263674
No robust homolog present in Genbank/EMBL as of 11/7/98
YTFKNPKKKNKKNKPNISIFLENTKHYDIFREGFVRDRHGLEASDWLLSTEITIRSL
KATNKHINACPLVWMTDDEKSN

CpN_0225 263946 264541
No robust homolog present in Genbank/EMBL as of 11/7/98
NSFTIKFLMLTKNAINSQTTTPQPNLTDAEPIASRAQCKSIIVISLFAFGMLLCLGII
LISIPILGLAAQVALGLIYSLILGIALANIGFLCLLLRCKQVPOKPTLPSESSKQSP
GSTPTALPQAGEFLEKQVVSATPILLPKNKDEELSAKVMKEGAEAAASSIKQAVLESTEK
LIDARKQESRREARKKI/AEEAEASRKRIQQQMAAQEALRKRKEEVAARK

CpN_0226 264545 264967
No robust homolog present in Genbank/EMBL as of 11/7/98
AIFNRKMPYANTLEFIQGTQSLCLPKFYGVFRHHYKQGLEIEDASHDWDFLEPPSTWK
RTLAAATPILGSLVIGLGRFLSIWSIREPQDSQEYKSIWHTLCAVLEILGLGIVALILKI
LATINAMPLGRVATFLFYS

CpN_0227 265467 265009
dsbB-Disulfide bond Oxidoreductase
KERFNIFVSCILKEIMMINFIRSYALPFAWASISAGTLISIFYSILNVEPCILCYOR
ICLFPLTVILGISAYREDSSIKLYLPQAVLGLGISIYQVFLQEIPLGMLDQICGRVSCST
KIFLFSYVTFPMASVAFGAIVCLLVLTKKYRG

CpN_0228 266242 265412
dsbG-Disulfide Bond Chaperone
VKDRADTFILNKEKFSCLKKENAFEFYFCSIKOLTNSSLRGLPNKKILVLTAMFFIV
CFGLYHKHTILPPKAHITPAKHFPTIGNPYAPINITVEEPSCSACAEFTTEVPFL
KHHYDTGIESFTLIPVCFIRGSKPAQALLCIYHHDQPADIDAYMEYFHRILTYKKE
GSHWATPEVLTKEAGLKINSRNVNPKGLBQCATSQGVNIGKNNLYGSOVGLQGLAT
PTAYGVGYLIEDPTFHETARIAQHIRQQAQVGDHDD

CpN_0229 266163 267560
Ct178 hypothetical protein
LNSKAFSLRIEQENFSFKFKKSLSTYNTANLTKSTFTFILLLLRKKDQGLRFMDKET
KENYHFRYRFLKLNILPAFLGLLLCSLNTLNTLQVDFVSDRLCSCLLIFLAISLT
KRSLWLGAAPLGIWTLFACVAGRSPTIFANDTLGIFAILAVVCISPTREALEVGTLP
EGFSYNSAGGRRAAVFLSLLGWLEARYLTASSIGITSSOSSNLLYSSIMTYSLV
VLSAGSERRWHTPKPIVATATALTGVILTLPLIILHQLRYDCWLCCLTIEPALAV
FAYDETATRLYISQFQDKRALTRASFGEYKHTLSWEERTVPLRKAYKQAFEGIS
FPIQQLAILVATVFEVNVNMSGLPTFRNFLNICWFIIVLFLAFAESLRHLRWMNLI
SAAILFSPVLFHIVESPMFLPIIVTGLILILSIGRRRTKRKL

CpN_0230 268277 267576
Ct179 hypothetical protein
RFFKALYMSQPLVTTSSSLRYVVLTEGEEKVACYKKAHNIWHGAPAIILAAALLMFC
IFGVLGSLILGALLEGASILYDVLFWLLPLLSVFLVLLPLNIYAYSHHKQVLAHER
ITQSNYKEIYDHCEKKTNNKALSILYIESQVLFPEYSKRFSSMLIGKTLKIPKIDSP
ESLKHLEIQALERAKENIYNNKQREKREAREAKNAKNTNPLWEGLOT

CpN_0231 268996 268253
taub-ABC Transport ATPase (Nitrate/Fe)
POAFVSIQDRGSMQAHRLCYSCNQVILKDSFQASPGTITILGSSGVGKTLFRLL
AGFLPLOGELLWNGSPLNRKDVAVMQKALLPWTALKNMTLSTELGINTSHALSNE
RLEEIHNFOLDGLDRYPDELGGQORALAAQCLSLKPIILLDEPSSLDVLLKEQL
YQDVALAKKENKTVLVTHDFHDVSCGLDVLVYIKNKTLPVPLDPSMRPLNGLCFIK
DLKHLVLT

CpN_0232 270134 269232
*similarity to 5'-Methylthioadenosine/S-Adenosylhomocysteine
Nucleosidase
KKFLMRFLFLILSSPLVAFSADNFTILEEKQSPLSRVSIIIFALPGVTPVSGDNCPIP
WFSHKKTLLEGQRIYSGDSFGKYFVVSALWPNKVSAAVACNMILKHRVDLILIGSCY
SRSQDSRFGSVLSKGYINDADVPFFERFEIPDIKSVFATSEVHREAILRGGEEFIS
THQKIEELLTHGYLKTTEHTLMGLVATGESFAMSRNYFLSLQKLYPEIHGDSV
SGAVSQVCEYSIPCLGVNILLPHPLESRNEDWKHLQSEASKIYMDTLKLSVKELCSS
H

CpN_0233 270439 270248
No robust homolog present in Genbank/EMBL as of 11/7/98
EKARTMFLGVLLFLRLISRSSVQETGIFHLETPLKIVLCAPVSTFIVVEMDVLKN
KQCS

CpN_0234 271246 270548
Ct181 hypothetical protein
FIMLOSCKALLSIVSILAFHPIPGMGVEAKSGFLGVKVGWFSKKEIQEERILPVKDS
LSWKRYDITSSSGFSVEFPGEPDHSGQIVVEPQSEITIRYDTYVTEHPDNTVYVSVWE
YKPEVDISRPENLQEGFSQMMQALPEQVLFMQARQIQGHKALEFNIIVCEVDYFRGMLI
SVNHTLYQVFMVYKNNKQALDKEYEAFSQSFKITIREPTIPSSVKKKVS

CpN_0235 271395 272177
kdsB-deoxyoctulononic Acid Synthetase
VFVRYLLMKPEESELCLGVLPARNWCRYPCKPLAKIHGKSLIORTYENASQSSLLDKI
VWATDDQHIIDHVDTPGGYAVMTSPGCTNCTERTGEVARKYFPAEIIVNIQDEPCLNS
EVVDALVQKLRSSPEALVTLVALTTEPEELTEKKVKVCFDSBCLALYFSPPIFLK
KATPVTLHIGVAFKREALFRYLOHCTPLSDAEDLEOLRFLHEQJIKHVCIVDAKSPSV
DYPEDIAKVEYITCENAYF

CpN_0236 272138 273766
pyrG-GTP Synthetase
KRTVHMPFKIFLTPGVVGLGKGLTAAGLALILERQRLNVAAMLKLDIYLVNDPCTMNP
FEHEEIVTDRVETDOLGHIHREZAAALHIGCATSGLYARVTKREREDYLGSTVQ
VHHTNETIQVILDAKELGPDVLTVEITATIGDIEGLFLEAIIQPRYDHSEDCNTH
MTVVPVLAQADEKPKQHEVQTLVGTGIIIDAILCDEKPLTOEVKSKISLPCNVNR
AVNPIVQKHITVEMHMLAQEKIAHFQKELKATVPENLDQWVILNQISQDLKPKVI
GVWYKIVQHRDAYKIFEAULTHAALRIJHAAREIIPDAEDENLTMELQVDAQLVCPXG

AKKMA:LNQVVCASPYATPOPVKEAWAVGSGPTDCV...LFESVGPDLVTSINGC
INIKKNAY:GZTIGAAKQALVQDIPCMALSDQNHISF...PEILKALVIYLLSQPP
CLTGLNINFTTPOGGSWEMRLVPPDEFFYEOPYL...KQYVYKISGVRIGCHP
SEELACMLNHSVSPFISQNSPGLMTLEEFQKTOENFNASLLSELTKIF

CpN_0263 296174 297136
yqfU hypothetical protein
DIALSRRLKRVPPSLAKYAFGRFMRSHGPRRTKFSFPLYSKTLWSFILGGFLAACGVO
MIALHRELI...
II INKKKGYTVQGIILFVNFIFALSGIVYKMMHTAFVSFLTYGIATKVMOMVILGLEDT
KSVTIITSSPRKLGHIMETLIGLTYIHAGGSGSEPRNLLVYVVERLQLSQLKEIVHR
EDPSAFIAIENLHEVINGRTT

CpN_0264 297730 297155
ubiD-Phenylacrylate Decarboxylase
MKRYVVGISGASGVILAVKLKELVNAHQVEVLIISPSGRKTLVYELGCSDFALFSEEN
LEYIHHSIQAISSSLASGSPVEATIIIPCSMTTVAASIGLADNLLRRVADVALKERR
PLILVPRETPLATHLENLLKLSKGATIFPPPMYFKPQSVEDLENALVGKILAYLNI
PSDLTKQWSNPE

CpN_0265 298632 297730
ubiA-Benzoate Octaphenyltransferase
KIIIVRLNYFLNLNFKYSIFSLFLSASTVFLASINEISQNSFKGFKISVFGAIAFV
PARTTIGVNVQCIDRFIDKNTRTSKRVLPAWLSLNFVWLSLFCFLFLFLCKLIRIF
SLGASLTLIMIVYPMKRVTFCHWGLGLVTVAILMNFCAFAESGLSMRLCLALLMGW
SVGMVIAANDIIYAIEDTEFREDEGLRSVPAYHGEKKAVEIAKVLNWSYLAIFYSGFVG
SLDKEFYFTAIIPLVILKVVVMYSNYSKQDGESEKFFLANIAIALSFLVSMTLFWSLS
R

CpN_0266 299181 299876
No robust homolog present in Genebank/EMBL as of 11/7/98
IMALDEINNNQNSQIASSTSQTSINQDRKTFACVTLLVATLMLISGIVLFTIGS
LGLSVPLSGILGTFAVTVGGTITGLTILVRKSLGIEQKNDLNLFLKIPSPARPLM
SKFSVTCSTTSIVLGMALLIGAVSVFFLTGYLQGLCAGLVGLGTALFVAGLARMSPRS
LAQDQSGSGADSQSNIVIGIEPKAAQEQKMYKMAVVRGEGDIPATRLTPEK

CpN_0267 300122 300910
No robust homolog present in Genebank/EMBL as of 11/7/98
VSLNKLNTNALLNQPEPAVCLANADPKYINQDRKTFACVTLLVATLMLITGTVILL
AMGSPGLSVLSTIIGTSVTTLTGALFIIGLVKLKSLAWIQYQYFQEVVKQYEPFS
IPKNDNVHKLTSCLPSPLDIESPSEASTPVSKLRIACSGVAIVLGVTLIGAVSVFFC
TGELALCVGFACLTALFVGLAGLRTHSLIAQIMYLYLYTSSALEERNETVQD
RNEINTYLTETCROQKREKALLE

CpN_0268 300914 301318
No robust homolog present in Genebank/EMBL as of 11/7/98
KQKALSLMSQCSSTSTWEMKSFVPMWKNPTPLSPISEDEFILAYEPFVLPKTDPE
NAQANPQGTSTPNVENGIDDLNPLLGQPNQNNANPGTSGNSPTSLPAPERLPETEENS
QEBEGSQNNEDLIG

CpN_0269 302468 301476
Dipeptidase
VAFRCVMTIDMHCDLLSHPHFRKDPAPVRCPEQLLGGVROQVCAIFVPHSGEPNCDK
QNSLFFSLPNQYPDIGLLSYEEENGSSQKSLSLIRSINENASALGDDTAPLGTLLAKL
IHLTKQGLAYLGIYVWKGDRFGGTEAPKRLSNDGKVLDDIMYELGVPIDLSHCSKLA
EDLSDYADKLPLNAVIAHSNFRSVLDHRRNLVDAHAKEIVRRKVGIGLNLVRSYVGS
LDLEKHVLAENLGLISSVLGSDFFYANEDENFFNECSSAEHPVLNQLIHRIFSGK
KAEISLSRAEKFLKQVIVQVNPKITDVKL

CpN_0270 303343 302468
ywlC-Sua5 Superfamily-related Protein
SIFGVIVPDKKAQITFLSPEVMSAIHOGKIVALPTDTVYGFVLSLYASEAERLYALKDR
EPSKAFALVYNSIEDIENISGVPLSTAKKLAQLFPAITLVKXRNPRFPKTLAFRIV
DHSVREIVDHCOTLIGTSANLSEFPASLTAQEIFADADHDLCIFDQPCSHGLESTVVA
SDPEIYREGLISRSVNIENIAGTEAKFIHRTSHAFSKHKIYTVKQEQQLVSFLSGSLDF
KQVMEHPKPNFYTLREALKKKTSFIVFYDINTSDYPELFPFLSPYIE

CpN_0271 303628 304362
Lysophospholipase esterase
KLMTDYSFRRKIGNIEAIECPGNPDPIIILCHGYGSLADNLTFFPSICSFSKLRPTWI
FNGILPLENDFRGSRACFPLNVLQLLQSLRYANGVGNLQEKYDELFDVLETPKEALE
ELILNLRNPYNEIIIGGFSQGAIALTHLVLTSONPYAGALIFAGARLNFQOGEGLKQCA
QVPFLQSHGYDEILPYHLGAHLNDLLTKLNGQFVSFHGGHEIPSVFQKMQVTVPMNI
DPARG

CpN_0272 305272 304340
dnaX-DNA Pol III Gamma and Tau
PNRQSDATYATVMHLEENQWEALLRKVYHQVPPAILLHGFTLPVLQDKAEQLESEI
LLSSSPGSEHKYQKIHPIQYFPEFGKRLHSIDLPRGKIKQIYISPEANYKIYIHE
ADMTLAAISAFLLKVFEEPPKHAIIITAKVQRLPKTIISRSLSIFIERGEKILCSKE
FSYLFRVACCEIPIVTEVSOIKESSETDKQVLRDKQVRFMEVLELRYDRYTNLGLKAS
ALNYPKRVKIELQLPLPLDKVLLIVESACRSANSSAASVLEWVAIQVLSQYKREL
VSVSPQQLSN

CpN_0273 305853 305227
tdk-Thymidylate Kinase
GGIVFVIEGEGGQKSSSLAKGDLVAQDRKVLTLREPGGLIGERLRLDIEPPHLE
LSRCECLFLFGSPAQHQIEVIIPALRDGYIVICERFHDSTIVYQGAEGLAGDFVADLC
SKVVGPTPFLPNFVLLDIPADIGLQKRRQKVFDFEKKPLSYHNRIREGFLSASADP
IRYLVLDAPEILACLDKVMHLQGLCT

CpN_0274 308368 305852
yjaA-DNA Gyrase Subunit A
ETIIMFKIDEIVPKNLEEMKEYSYLRSMSEVITSPALPDIDKIKYVORRVLYAMKOL
ELGIAKHKKCAICGDTSCDYHPIGEGVYPTLVPMQNNAMRYPLVQGNFGSDQD
PAAAMPYTEAKLTHIAMYLMEDDOKDQVIVPNYDEKHPVVPFKNLLNCSSGIA
VIMATIIIMHNLIELIENALLLANQASVDEILQVMFQDPYQGLCCGEGIRSAVTT
GRCKIKVPAKLIVERNEDKHREIITTEPMYNNKSPLEIQAIALNNEKTLACISDVRDE
SLDKLIRVLEIKKQSGEIIINRLYKTFQVQVTCFAHMLADKYLPRTHSHRMSIAFI
RIKKEVIPPRTYELAKAETRAHVLEGYKALSCLDALVLTIPESGNKEHAKERIIESG

FTEPOALALELRLYGLT...OKEYEELLNKIAYYKQVLSDEGLVKDIIRNELQDL
LKHKKHVAARTTIEFDADDI...ITNEJVIITGGDDYVVRMPVKVFEQRGGHGVT
GDMKKAGAGFLKAVYGAFTKDTLLIFTNPOCYVWLVQWPEGERRAKAKPIINFLBGR
POEELAAIINIKNFDAAGFLATKRCVVKVSLDAFNSNPKRGIALEIDEGGDLIAAC
HIVSDEEKVHLFTLHGMVRFPHKEKVRPMGRATVGRVGSVKNNEKVKVSCQIVTENQ
LIVDCQDQKGRSLVEDFRETNRGQGVRSILINERNGVGLAIPVTDHDSILLMSQCGA
IRINMQDVRVMGRSTQYRLVHLKEGDAVLSMEKLSNENDDVLSGSEECSDTVSLR

CpN_0275 310749 308772
yjaB-DNA Gyrase Subunit B
HMDPKENNYDASAITVLEGLAVRERPMYIGDTITGLHILVTEVVNISIDEAMAGTCS
RIDVRILEDDGIVIVDNGRGIPIEVHERESAKQGREVSALVVLVLAGGCKFDKDSYKV
SGGLHGVGVSCVNALEKLVATVFKDKCYQMEFSRGIIVTPLQYVSVSDRQGTIEIVFP
DPKIFSTCTFDRSLMLKRLRELAFLNRGITIVFEDDRVSVFQVTFYEGGQISFVSYLN
QNKESLFSEPIYICGTRVGDDGEIEFEALQWNSGVSELVSYANNIPTRQGGTHLTGFS
TALTRVINTYIKAHNLAKNNKLATGEDIREGLTAVSVKVPNPFQGTQKQLGNSDVS
SVAGQVVGAEALTIFFEENPQIARMIVDKVFVAAQEAARKARELTLRKSALDSARLPK
LIDCLEKDEPKEMYIEGDSAGGSAGKGRDRRFAIIPIRKILNVEKARLQKIFONOE
IGTIIAALCGIGADNFNLKLYRRRIITMDADVGSHTLTLTFFRYHALLIENEC
VYIAQPLLYVSKKDFRYLSEKEMDSYLLMLITNESSILFKSTERELRGEALSFINV
ILDVESFINTLEKKAIPFSEFLMYKEGIGYFZYLAPOKMQGQGRYLSDEEKEALAQ
EETHKPKIILEYKAVFVDIQNLKEYGLDLSYILIPQKNEIVIGNEDSPSCYCLE
EVINYLKLRGKGIIEQRYKGLGEMNADQVLTMTNPEQRTLHVSLKDAVEADHIFTML
MGEEVPPRRFIESHALSIRINLMDI

CpN_0276 311140 310793
CT191 hypothetical protein
DMFLKRRKRGGSQVONKRTASPIKHAHYLHNYLQELQKIMAAHPHDAIDAWNQVFRDKY
KQMSQAIFGRDILLVYVNSLYALIKQTPQNDLMSLYQVASHVQIREIQFLG

CpN_0277 312003 311404
No robust homolog present in Genebank/EMBL as of 11/7/98
NISIFYPKYFIEGKEVLKLNLPPIFYGVILMIINVRPAFGITSVQGFSTNFQAAIPIL
NIVIGCSRISSYAEIDEEVAQKLEKSLKSTVNLHARVRCVVAIGAGIACVWVYVYKGFCSQGELE
PIEVKTLISPDKPYPTVYV

CpN_0278 312884 312060
*conserved outer membrane lipoprotein
RDSMKKLSLLVGLIFVLSCHKEDAKNIRIVASPTPHAELESLEEAKDLGKILKIL
PVDVIRINRLLDKQVANYFOHQAFLDDECEYCKGLDVVIAKVLHAGQAIYSKHS
SLERLKSQKLTIAIPYDRTAQARLHLEEGCLIVCKGPNANLMTAKDVCKENRSINI
LEVASAPLVGSLPVDVAIPGNFAIANLSPPKSLDLEDLSVSKYTNLWVIRSEVDGS
PKMILKQLFQSPSVHFFDTKYGNLITMTQDNG

CpN_0279 313546 312875
* Possible ABC Transporter Permease Protein
KKIMOSDLIQLKETVNTLYMVSTAFSCAIGMLGLFCTSPKSLNPKKSLYATIS
MILESFAIPRAILLIVLPITRMIWGTSLGPTASIVPLTIGAIFFVTVIVDAFRNSAL
NYLSALVAGCPKRAILFGLIPESYQALIFSLKSLVHILISCTLAGVGGGLGQLLL
QYGYRREWSITTSVITVLVLESVIRLGDGFWRRVLYKRGIL

CpN_0280 314593 313550
dppF-Dipeptide Transporter ATPase
IKGEAMVSPQHSFIIISVDQVSKKLGHDILLKSVFSVYSGEVFGIVHSGSGKTLRLC
LOFLDMPTSGISVAGFNSLPTQKFSRRNFSSKVAIISQNYGLFSSKTVEINAYPLRI
HSEMEKSEVSEVQVYDILNLFNLVYHRHDYAGNLGSGQKQVATARAIVCQPEVWLCDEI
TSALVPTKSTENIERLLQNLQERGITLVLSVSHEDVVKICSHVLMHOGAVEELGTTEE
LFLASSENSITNLFHEDINIAALSSCYFAEDREEVRLNLFKSLAQIGISKVITQGLVS
INLSGNINLFRKSPMGFIIVLGEVEQRKAKKELLIELGVVKEFY

CpN_0281 315033 316103
yhaA-Predicted 1,6-Fructose Biphosphate Aldolase (dehydrin family)
ISLRHRLMLNIHIDILGNDENLLSYQCKHITDKDLTPSHDFVDFKGLSDRNNRVLRS
LQTMFPHSHGLRANGYLSILPVDQGEHSGASFAINPIYFDPENIVKLAIESGCSAVAST
YGTLSLRKRYAKHIFPMLKLNHNLNPTKYHQIIFPTQVEAAYSMAVAVAGATVYFQS
ETSNEEIVAVSNAFKARSGLATVWLYLNPALFVANGVDYHTAADTLQDADHLGATG
ADIVKLPCTCGGKFAINFKGTDERVYSELSSNIPDLRCRYVLSNYSYGVGLINSQGP
SGKNDTEAARTAVINKRAGMGLIGRKAQORPLSEGIQLLLNLVQDYLDPNITIA

CpN_0282 316084 317529
xasA/gadC-Amino Acid Transporter
ILILQSLNFSKVKFMHSHSKPTKPLGTFTVGMLSLAVVILSRNLPLTAKHGLSTLFFYGL
AVICFMIPYALISAEASFKPGIYIWARDALGKMGFFAIIWQWPHNMTWYPAVLPIA
STIVYKINPELANHNYIATVILAGFWITLTFNGLITSSALFSSICVICTLIPGVILV
SLALFVIFSGNPIAISLWGNLLPNFSNVSSVLACMLLALCLEANANLSDMWNPRK
NYKAVFIGATATLTIVLQSLSLAIVIPKEEISLVSLG/RTFTLPFDKYNLGMWGTGIV
VMTIAGSCELNAMFAGTKGLEISTQNDCLPLFKKVIKSNVPTNMLFQGIIVTIFTL
LFLCLDSADLVYWLTLALSVMYLAICLFLAGPLRIIEKPAQRLYVSGPKFLGICTM
SILGILSCAFALWVSLPPRELAQISEGSIKYTTFLLLAFSLNCLIPFGIYFTHKRLSK
KS

CpN_0283 318581 317532
No robust homolog present in Genebank/EMBL as of 11/7/98
GRLGLYQDLIKNAVAKIISFRKSPNPVKLLIKFAKDKLKNSSIAPLYEVLLLEIAPG
EELILEVLSLDPMLKSLMPLKHKHTLGIEISSETAETIESCSGLISINLLSGCLCRS
SHDRCAQVNIIOQFCQFSSEEVONFVEORNILTFPLGEGDEVALNLQGLRLDLIV
PNALYFEPDPCWQINSECAKDAEDQEDFNKTKKACQEGGLKLVLPALGITSIPOLL
RARRFQQAELMAIPRKKMKQNPFIIEALLLEDEEFTGVKYLKLMITHWDLKLIHA
LYGYPQGLICQIETETFRANINPEAFQAAIQOQRIILGFLPKMLLD

CpN_0284 319054 318551
No robust homolog present in Genebank/EMBL as of 11/7/98
FLIMTIPARVVPVINEVANTSGGGLKGLRLDITVILALAIATLMELVYVNTIES
WTFVIMLFLVWAVAVLVEYQGGIETKTVGCTSGWFFQDNLNLLREEDS
VSAIDELLKNEPAPKRRKMLPYNILDEQYKINSEPEEDHTKIL

CpN_0285 320601 319051
No robust homolog present in Genebank/EMBL as of 11/7/98
KELFLPFFFTANKETTAHELTYRQCELEPITGLCLALVLLLAIVVAAKATVIL

AAPGLLVWGCAASVCSMAIVSLMCLYKGGKPLIEP... OPTKDEIKDPESLKPV
PVEQSLPKERKTVSFKAIPSIIVEDDFKPYVQSTF... /SKPIAERMSQLEKEIT
TLIVDPRALEESGSKSGSLRGVISEIKNLFPLP... /SLTACLPLRGSIVEEYA
SSDLLILLTKPEPLAMVTOQLIAHNSLKTEKRKLTPHMOKLVLISINFWFYGWSLEEKC
IEKIVAYDPLLDELKAHLEAGNIVQFLSFSQSEMQREFRALFSDAQELPSAKDGSN
YVPAINSSEYMYDKDLSVGLKSLGERLAFCEKIPSPSSMNTTSVASHYKDFSLITFTF
SNQSVILQNPFLLELILHNPKQCTFLKCLLEKAMPMSNWAALFRPMLMGLMCSGIARK
KELKIIAEHLVVPFKETTOAIASGKILDLQLHDF

CpN_0288 322149 322149 322149 322149
mgE-Mg++ Transporter (CBS Domain)
SCRESKGMVGEQNRNEELKLPDAPSSGNLMSDRTSHLDELSPFKLAKFTCLSTDIHSH
DLKSLVILEYNPIDLAYAVSLCLPSESRAILYKNSLCTAKVAFIINTDSASRSLSELRSD
SEVCALEIQMPDPAEAVWDDIPDRRYRRIELIDSKKALKIRDLQKHGRTAGRLMTE
FFAFLEMTTVKDVASACIRSNPGIDLTRLVFLDFKGELOGVVTRSLIINPPMSLQKIM
NQIEHKVLPDAREEVVDLVERYKIAALPVDEENFLGAIYTVEDVVEAIEDIADETIAR
MAGTTEDGYOTCHVQVORFLVFLVTLFAGLISASVMAYFOKIPALLALIIFFPL
INGMSGNVQVCSITLVRSMATGTLFSGRRRETIFKEMSIGLLTGVLGILGGLVYVLMG
FLGLNIFSGGGIQLGVTVATGVASLTATTLGVLSPPFAKLGVDPALASGPVITLAND
IMSMIIFFLIAGGINFLFFN

CpN_0287 324230 322089
No robust homolog present in Genebank/EMBL as of 11/7/98
RCHMIRSPPLPFISSKRALNMLGLQDEPSCPDDVDFLSEIEELAAQDEPSEGYALRS
LLMTTHNHPKVRKVIYGVSYGLKHSMSIFIDVLTVIDFLKGLSASDRSLCSAR
TCINFLYSQTEGMKFLSEVDFNRLILQLKMHPLQKRLGWEHFRIGAKGEEVLSAS
ASVYQAVGSRFIELYHKLHLESDLACGMKCLALALDLSNNNAIHADYAKGLVVLGTROG
KSLILIEGMEHFSKAIPLFSFSDGDTLAYQNYRYALASVVLFDLTYKKEHFDQAMNLT
YQTVQAFPNLSGLMWMGELLIRSGWLSNMKYIEVLEKLSLQKKTNDPILASGLLAT
GAILGLVYLEENPLFKDSRHRLISAMRTFPGNSALVHALGVVQLSALYFNDSHGLFA
SCFQSCLEWDLATGMWQKLFDAFYSWGIKKKSARLLKAVDVASRLSLRPEAFFLWSD
RGLALKCLAEATIDEAYKEIFLSESLHYORAWDLSCRLIELEWQSHYLLAEGLFV
HYDEAYTLLTKVDLTLSSRVKLILAAVLGLKGRLLQDTPAEAREILEPLVEVLEDE
NFLLEGLKYFLFLWKNKNGLGLARTYLEKATSLGCPPEAYTLGKFYAVIKDVKWANG
MVIRSAQYGVRITEAKWLPYLANLREIHAFREVENQKGRMLGNKTEMRKN

CpN_0288 325785 324571
CT288 hypothetical protein
ISITIREFLFFGFECAKAFYNVMSFNLSTNESLRPISPKASFPKQGWQSYFRSLARK
HRSDTILSVSVCKVKNYDANLFLVLTIVIALAVVGLILFSLIASIQTGLVITSWPLVTAA
ILFPIITLLTGGMYILHRLGKQVVDISGVCIIPFSRRCWPISSHTLEKPEKHVSACS
LDISITLSADGSGIAAVYQCPPLFLFRAFCFPCGICPMCFVALLRMITNLRFLVVPYIIF
RMVYEHFFCKHLPEDDRFIYKDVAREMGRSLAAFLKAPFYASACMIGAFYSLLDPLAGRV
LMGSEVMDNDNVILARVSLANEALSLRFEFGGGRKGLGQHAFLVLMCCQPSVFLFD
KGETVSGAHPISQLPERGLDTSGRYPHISVIPSNDGNSAKNFIV

CpN_0289 325797 326996
CT289 hypothetical protein
NENRMLKKORSHYKKNLILLSSILVGLGSGVSPWIVYSAECIANTFLKFLRLLSIPL
VBSGLSTIISQNFNTMTLGRKILYLLTTLTVAASIGLLFLFLRPMQITQDALATT
TKGNPLGYLDVLDLTPENIFKFFPLQGNVISAACLAIVLLTASLFLQEKKEHFNVOFNS
FESIFLNLARGGLKLLPAMLGPSVILFKELKQSNLMTFAEYLLCVIGANLAQGFVILP
ILGKINKVPLKAKAMKSVITAFKSSAATLPLTMEAEEDLKINKNLSRFPFLCS
VIRMGCAAFILITVLFVATSNMISLPSMLGWIFLATLAAGNAGVPMGCVFTLLSL
TSMNPLSILGLILPFTVIDMETSILNVSDCCVSLAN

CpN_0290 327027 328523
Na-dependent Transporter
RSALTMNKKHAFSSRLGFIPIFMIGIAGVAGNIWRFPVAAQNGGAFILWLCFLFLWS
IELIIIEIGKLTAKIPALIKTAGKKFAWAGGFITLVTTCILAYSTIVGWLSYFI
YANSGKILHNGDAFLWTSYHQSSIPLWHLTSLGLAYLVIRKGIVHGEKKNILIPAF
FLCTALLRAVTLPGAQVIGIKLFCSDKSCFNNYKWIIEALTONAWDTGAGWGLLVYA
GFSKKTGVVNGALATCINNVLISIMGIIFSTCASLDLGTTLQDQAGASSIGTIFI
YLDELFTLPGGIYLLTFLSSIFFLAFMAALSSMISMLFLLSOTLAEFKIPKISITLA
TTEAFVLGIPALSLSLTFSSNQDVTWGVVALVNGLFIYAAALVYGFPLKKEVINAPGDL
RLNAFADYIIKYLPIEGILLLGWVYFEGLPENQWNNPISLISLGLSVLWLSLIL
WKNKQLYLRFSSRYNHEIL

CpN_0291 328658 329194
IncB-Inclusion Membrane Protein B
EKHMSAPIPTQELSDQITCLNVQVQVSELAENKGDIEGLKTLTAALDAGIQPSAD
EYISQCTAAALILSASEKPGSGPSGTEGTVQSPCKFKKVLAVLVTIIILIAIAVLIA
CIIAACQGFPLLSALNLYITGACVSLPIASTSVKICLCTGVANSLIPVITRTIA

CpN_0292 329201 329836
IncC-Inclusion Membrane Protein C
VKNTKNSDFMTSPIPFQSGDASFLAEQPOQLPSTSESQVLVTOLLTKMKHTQALSETVLO
QQRDLPTASIIILQVGAPTCGAGAPFQPGADDDHHPIPPVPVPAIETETITRSELO
LMRSTLQOSTKGARTGVLVVITAILMSLAIILIIILAVLGTFTGLVQVALLMQSETNLI
WAMVSGSICFIALICTLGLILTNKNTPLPAS

CpN_0293 329940 332723
CT234 hypothetical protein
VWSMORVRLLLFNLHGEKRAFLFFLLGLVWGICGYCTLSLAELGFIKGLSALPKIY
LCCSLILCVLSSLLILYNLFKHSIATLFLIPVLSILCNFYLLSSIFAIDPPRSLFF
YRIVWLSLITLSYTSFWGVDFQFNLDQGRKHFIENAIPLGDAIGSGIISLWVITIGI
OGILLITLTAALVLTPIVYVSKLSKLSDDHDLFDTCAPPPLSKALKCFYDKYTFYL
LCFYFLMQLLAIAATEPNYKLIKEIFQASKEEFELVAHIZKCSLWISLGNMCFALFAYSRI
VKRLOVNIITLFAFLFLSLFLPFTKTLTSLAVLWVREGVTVYALDDNNLQLLIYGVN
KIRIRNIRIVVESIEPIGMLVWGLICFLISQOYVPLIISLAIILVCLVRVYIYKVL
KNLGAQALQTRSMQDNKIMTVKQKQVELFLLAHLKHPSEHQTAFQHLNLSASRV
LPDLIAIMNKLSPNLKLTINMVKSSLWAKDLTLKLRWTSIFPHPAICAHLYFAE
HDLHITHIAEDLYTVGRDLALTLVTRQEAQPYRDLADKRLKELLNSDOPEDIVMG
LTILKLEKNQNFPIILDFLNTKNEIDLVTCKLITSVRANIKIYCFELLRQROCHSN
DEAQQYLLKTIHALDIPVKDLMTTQKLNKTRKYAEAMIGLKEVAPAFQVLTDE
THNRCKILAAKALCKIDNWLKKHIAKYLIVKSAKALFYSGHYIKKYPYTNLJLA
NFIENNYAVNFMPLSLILGSSMEGSLVIALTSKKNOKIKQALESLEKIVDSHLFSL
LEPYNQYKMYSEKYVFKCVIPLTKELLMMENSPCLNKNITAOQLKEELGYCOPDF
QVNTVYNQHHDFRTESETLISFLSI

CpN_0294 331077 333502

CAMP-Dependent Protein Kinase Regulatory Subunit
IRNFFMNLIDRAFLLLKTI... DMILLITADKETTIFKIKGNVETIQGPGFSFYII
VEGYITISKEKLESPLNKLPLDGEELFNKPREYNAGANTQVRMLVLSKQGLNIVE
ECPVALGFLLEYAKQIKREP

CpN_0295 333866 333627
accP-Acyl Carrier Protein
AMSLEDDVIAIVEQLQVDPKEVNESSFIEDLNADSLDTELIMTLEEKFAFEISEEDA
HPLTPIVNTVNTVYKRAEQ

CpN_0296 334790 334822
CT296 hypothetical protein
KIPIRGMICMDITLVGKKIVTVGSGRGIGLVKFLENGADVEIMGLNEERGQAVIESL
TGLGGEVFSFARVDVSHNGGVKQVQKFLDKHNKIDILVNNAGITRDNLLMRSEDWQSV
ISTNLTSLYITCSSLVIRHMTARSQSIINVASIVAKIGSAGOTNYAAAKAGIIFTKSLA
KEVAQNRINVLNCLAPGFIETDMTSVLNLDNAEWLKSIPGRAGTPEDVARVALFLASQL
SYMTAQTLLVVDGLTY

CpN_0297 335724 334774
fabD-Malonyl Acyl Carrier Transacylase
SRSNKNDGMKKRYASLFPQGGQSVQVMGQDLYMEYPEVRELFDFANERLQFLSTSIMFE
GPEDLLMETVHSQALYLSHMAVVKVLSQRSSITQPSVLGSLGEYVALVASDRISVLVG
LELVKKRQQLMNEANQSPGAMAALLGLPEVEIBENTISLGGGIVANNYAPKQVLVAGI
AEKVQAIELFRDICKKAVRLKVSQAFPTPLMQVAQDGLADPIALCMKDDSLPLVSHV
VGLSVNTEEMRECLARQMTSPMLWQSCYHIESEVDEFELEGPVKVLAGLNRSIGISKP
ITSGLTQAIKELFSEV

CpN_0298 336742 335717
fabH-Oxoacyl Carrier Protein Synthase III
YTSFFLYVSESLYVSNMKNKAAIWAQTSGLYPEKVLNADLEKMDVTSDEWIVTRTGKERRIA
GPOEYTSILGAIIEAKAIENAGLSKQIDICIIIFSTAAPDYIFPSSGALAAHGLIEDVPT
FDCAQCTGYLYGLSVAKAYVESGTYNHLLIAADKLSSVFVDYTRNCTVLFGDGGAAVC
IGESRPGSLINRLSLGADGKLGLLSLPAGGSRPASKETLQSGKHFIAMEGEVFKHA
VRMETAAKHSIALAGIQEEDIMFVPHQANRIIDALAKRFEIDESRVFSKHYKNTA
ASSVGTALDELVHTESIKLDDYLLVAFGGGLSWGAVVLKQV

CpN_0299 336726 337415
recR-Recombination Protein
RKLIVVYSESLYVSNLNGRPECKNKHITMTRYPDYLSKLIFFLRKLPGIGFKTAELKA
FELISWDSBOLKILGNAPHNVASERSHCPLCTFLKESKADCHFCREERDNQSLCIVAS
KQVFFLESKVPKGRYVGLSLLSPITGKLIENRERLSIKSRIETLCPKEIILADITALE
GDATAFLKQELQHFVNISRLAIGLPLIGLSFDVYDSGTLARAFSGRHSY

CpN_0300 337768 340152
yaeT-Omp85 Analog
GRLGLMLIRNKNVILQISILALIQTPLTFLSTEKVKEGHVVVDISIITEGENASKNHPL
PKLKRSGALFSQDFDEDLRIKAYEYDSEVPEKVEFSEKNTIALHLIAKPSIRNIHISG
NOVPEHKLITLQIYRNDLFEKFLKGLDRLTYTYLKRGFYASSVDYSLHEQKQKHI
DLVLIKINEGPGKIKQLTFSGISREKSDIYQTKHSTTTSMFTGAGLVHPDIVEQD
SLAITNYLNNGYADAVNNSHYLDLQKGNILLYNDIDRGSRYTGLGHVHQGFVPLKRLI
EKOSQVGNPDLYCPDKTWDGAKIKQYAKGYINTNVLDVLFIPHATRPIDYDVEYSESG
SPYKVLKIGITGNHTKSDVILHETSLFPGDTNRLKDELTEQRRLNTGYGQSVSYTVR
SGLPDMGADQYRDIIVEVETKTGNLGLFGFSSLDNLFQGIELSESNDLFGARNIFS
QRFCRLRGQGEHLFLKANFGDKVDYTLKWKTKPHLNTPWILGIELDKSINRLSKDYAV
QTYGGNVSTYILNEHLKYLFGYRGSQTSLEKRRKFLGPNIDSNKGFVSAAGVNLNYS
VDSPTPTIRGGVTFEVSGLGGTYHFLKLSLNSISYKRLTRKGLIKKIGEAQFIKPY
NTAEQVPSERFVLGGTETVYGYKSFIIIGPKYSATPEQGLSLLISEEFQYPLIRQPN
ISAFVLDGFPVGLQYKISLKLDRSSAGTGLRDMNVNVPVMLGFGWFPFRTETLNGEK
IDVSRFFFLGGMF

CpN_0301 340163 340762
(OmpH-Like Outer Membrane Protein)
IKDLSKEIFVVRKGFYWPFSIPKLVOVMKLLFSTLLVLGDSAAHANGLGVNLRKRC
LEESDLGKETEELAMKQFVKNAEKEIEELTSIYNKLQEDYMESLSDASEELRKFK
EDLSGYNAYOSQYQSQINQSNVKRIQKLIQEVKIAAESVRSKEKLEAILNEEAVLAIAP
GTDKTEIILNESFKQKQ

CpN_0302 340766 341866
lpxD-UDP Glucosamine N-Acyltransferase
SKNFKPSEAPVYTLKQALQLQVEQVGNITETPISGVEDISQAQPHHIAFLDNEKYSF
LKTIRKAGAIILSRQAMQHAHLKKNFLITNESPTLTFQCIELFIEPTVSGFGIHTPTAV
IHTPATIEKNTIETPVVISOAHIGSDTYIGAGSVIGAHSGVLGANTLHPKVVIRERVL
MGNRVVQPGAVLGCSCGFGYITNAGHKKPLKHLGVYVGDVDEIGANTLIDRGRFKNTV
IHETKIDNQVQVHHVEIGKHSIIVAGQAGIAGSTKIGEHVIGGGTGITGHISIAHDVI
MIAQTGVTKSITSPGIYGGAPARPYQETHRIAKIRNLNPKTEERLSKLEKQVRDLTSPSL
AETPSEI

CpN_0303 342982 341921
CT303 hypothetical protein
REQKGLHMDVSRKINRHTQYVDSIDGVINKFDHFKPSEDKSRDHEELEKLLTITKRIV
ASAQEQFNKRTDSKNYLLKTKTOLPFFKNEEQTKELFAMLTSMOKKIAQLFYSPQCS
DWVETEVICHNLDSIGLQVGLLCCGLFEQCEHVVTNKKLDLPLLLGTTVNSLRYLL
TYRNSILLNCQSSSELGKGLVLDKQHGVAFTLIFKEITVDLILNLYVKLIQGLKRSQNIQ
ARIYNDVPTLPSVSSSIALYSLANTIRGLALRGLVDSFSLKFISSPLNSTHTAKALN
SGGECFISNLDFNLQMKIVMQLLRGTGIPETILNKNIMKILMKRRVRSLYI

CpN_0304 344091 344154
pdhA/odpA-Pyruvate Dehydrogenase Alpha
DQKPLKPLPYKVMDSAPYNIAIQTEKSTVERILDLYGPASCIKFLKOMVLIREFA
RGEAYLEGLVGVGYHYAQEAVATAAIAQTLDPWFSGSYPCIALAILNIPLOEIAA
ELLKETYCALRGCSMMHGNPFYKGFVWVQIPLAAGAAFTIKYQKQNRVSLCPI
IGAVAYVHETLNFVLIQILPLMLIENNAWMTSILNRAVAKQIAEQSGSYDRAV
TVNDEHILAFREAYRYVDTESVIVVEELKTSRFRGIEIDNIVYKREKMOCLFKK
DILVLAFTMLIRLEVLTREEFPIRQETAVLEAFNARLQSDPVTTLTEFIVYA

CpN_0305 344142 345117
phbB/odpB-Pyruvate Dehydrogenase Beta
KKECMIPKIKLILREALREAFREKMDPKNVILISFEEVDTYVAVKVTYGLLDKWKPKRV
IAPILCAAFSGITGAAISLPLIETEMWDEPVALQVILCHAAKMPMDIKFVPI
VIRGTHKAAQVQHSKVEEVALEATLITLAIHSDYDAWLEKATIRNNNPLFLEN
ELEYNLKEVITEEYVLPKKAHRYVYHILPITPYHMYDKKATKAAKFWILTEI

[illegible]

AEGLTQCDLNFVKVAEYHKI (HDKFPTAGKVEREV) VDSVTQVLNLFSSALR
QTSSRLFGGADKROQLJAMIANALDAVINNEDYPKAS KPSW

CpN_0325 370666 371148
CT325 hypothetical protein
KRIAMQNYEQLESAPLANTTLPADKNNKSLIRFSDTHVPVQIEEDGNSGDLAVSTLL
GTLPEVYFRERIFKALGNSVGPSSIKGILGVEVTOQLYLSILSMNYLNGEKLFYEL
KLFSHLAKIMWESLRTNLPDHLVGLTYVA

CpN_0326 371118 371448
mal0-Glucanotransferase
PSCFGNLLRRVNLVYTKSPHAWKLGTSKPHGIYLPFSIHTKNSCGIEGFDLILP
LISWCKQKQFSVLIQLPLNDTGEDTSYNSISSVALNPLFLSLSSLPNIDTIEPAVKLQ
ARFRFPSPVSSYTOVKEKWAFLREYQKCKSSLEGNSNFSEFLESERYWLYPGTFR
AIKHHMHCEPENNPKSLTDQENFDLTKKFHDVFFSYLQFLCYQQLCEVKAADQHH
VLLKGDPLILISKSDCDWYFRDYSSSSSVGAPPDLYNSEGQNHPLIYNFSQAKDDY
IHWKERLYAQNFSYVRLDHIIGFRLWIDWSSGRGRFIPDNPKDYIKQGTIELSTMLG
ASSMLPTGEDLGIPODVKTTLTHLGCIGTIRPRWERNWESDAFIPKLDYNPLSVTLLS
THDSDTFAQWNLSPKAEKQFAKFLHPQKLTETQIDILKLSHESAFIHNFLNFDY
LALCPDLVSKVQLQRERINTPGTISKQWYSVRPSLEELAHHKFKNGYIEKILTGL

CpN_0327 372927 373211
r128-L28 Ribosomal Protein
RIHKNNMSRCKPLTGKPRRGYSYTLRGIAKKKKIGLKVTKTKRRFFPNMLTKRLWST
EENRFLKLIKISASALRHIDKLGLEKVLERAQSKNF

CpN_0328 373220 374992
CT085 hypothetical protein
LKYREIFMSFLRRHISLFSQQLIDVFPVSPNLELAEIHRVRIEDQGFALLFHNVIGS
SFFVLTLNLTGTHRRVQDQFSQAPDNLTARVAHLISSTPKLSSLMKSRDLKRISSGLKK
ARFRFPSPVSSYTOVKEKWAFLREYQKCKSSLEGNSNFSEFLESERYWLYPGTFR
NTMGLHFOIQKGGMHLYEAEQKQNLVPSVFLSGNPLTLTSAIAPLPENVESELLFATFL
QGAKLLYKTNNDHPLHLLYDAEFILVGSPPAGKRRPDPFGDFHYSSLQHDFFEFCHK
IYHRKDAIYPAVVGKPYQEDFYIGNKLEYLSPFLVPMGVRRLKSYGESGFALTA
VVKERYWRESLTALRILGEGQLSLTKFLMVTQDEVLDRFSVLETILERLQDRLDII
FSETANDTLDYTPSLNKGSKGIFMGIGKAIKIDLPFGYQGGKIHGVQDIAFPKRGCLVLE
TSLDRCIKSLHHPDLKSLILILANRDIQSEKDFLWRTFTRCAPANDLHALSHF
ATHRPNNYFPFVIDALMKPSYPKEVEDPSTKQKVSERWHAFFPKETFYI

CpN_0329 375085 376146
Phospholipase D Superfamily (leader (33) peptide)
KMMKQKDKLRCVITSTLILVGIAPARAGDTFKTFLKSEAIISYNSQCNEDMRKILCD
AIEHADEEFLRIYNLSEPKIQSSTLROAQAKNVTIYQKFKIPQILKQASNVTLVEOP
PAGRLKMHQKALSIDKDAWLGSAANTLSRLDNLILCMHSELCDLITNTSGDFS
KQDTGKYVLPQDRKIAIAQVLEKIQTAQKTIQVAFALTSEIIOALHOAKQRIHVDI
IEBSSKLTFFQLRQNLNKKDFISINTAPCTLHKFAVIDNKTLAGSINWSKGRFSLN
DESLIILENLTKQONKRLMIWKLAKHSEHTVDDDEKEIIEKSLPVEEQEAA

CpN_0330 376930 376202
CT083 hypothetical protein
FISIEMLLSRQLFSLPSRFQDLHVYRKESLKLQFMTMVGIEVIVLAEIKEEDLR
RKEPVRKREKYNLRIFRVLSRFDVRIIRFDYPGALSQSTAKDSRONSPLEKISEE
ATNEAIRLALLAIGDREQEKKQRHRYKLLGQKQAKVLSQLRHVHLDFKLYCDSKKK
DQEKDEKQKRSIKVTKKKKISLGAASQAIAAAEAWIARNKGVLETAFTLYQKD
EBA

CpN_0331 378452 376701
CT082 hypothetical protein
IQRIIMAVSGGGVQSSDPGKWNPAEQEQAEGSPKLESIFSETQASSAAKQELVR
SSTQMYATESQINKAKYRAQKQNRISPEMGFFLDASGMSGSSDISQLSEAKKSA
KASDSGGTSSLLPTMDVALLKQNRISPEMGFFLDASGMSGSSDISQLSEAKKSA
FSGARSLSSSESSSVASFGSQKAIEMPESEKVNATVRLGEMVSSLLDPVETSS
LVHIAEMTLNEGMDLSDQGEVESTAMTSRAVEGKVKVSSSDSPCANPTGINSMTLE
RAEKAEQKESREQLSEDQMMLARAMAGLLTGAPQEVLSNSVMSGSPVTFPPKFSOTL
PTGSGDKSKHKSPIGIEIRNFTNPSPLRGCTVKSAAVKSLLPHSPMYRFPKQSVSREE
PEAVVKESTAFKNPENSQNFPLIAVESVFPKESGTGALGSDAVSSSYHFAQRGVSLL
APLPRATDYEKLEAHKGGPGGPPDPLIYQYRNVAVEPPILVLRSPQPFSSGSLRSVQGP
EASVHDDGGGNSGGSGDQRRGSSQKASQKQKGLKSTDI

CpN_0332 378676 378536
CHLTR T2 Protein
YLDNRIRVILARQRTLLHLLAVLCPPISFPTQGVSPVFCFLDF

CpN_0333 379117 378800
ltuB
VDFVFVFVFMCKPKKSRDRALAEIQKSTEVLLKPKARIKAKRRKFLIAKEQKTLKHR
AQEYDQLVRLSLDSQKQDTPKVLIFNYENGFPVTDKDHFSKYIRL

CpN_0334 379308 379823
CT079 similarity
TMSVHITPRKCFILCLSMFTLPTLFPKHLILFSPYIVLCFYCFKDKGLVLALGCGVL
SDALGSGRGVFLLLYPLTALITHKAHLIFSKESKAAVYVNMIFGVFLLTIPMCALFG
HEVRWSIDVLMIPKCSFLDNLIFTSVIYLPCAINSCQHKMISFFRRLVCY

CpN_0335 379808 380674
tdiD-Methylene Tetrahydrofolate Dehydrogenase
EIQMLPPIPAAEKILQRLKEETSSQPSPLAVLIGNDPASEVYVGMVKMKATEIGII
SKAKHLPODSTLSCVLKLIERNQDPSIHGILVQZPLPKHLDSEVILQASIPKDDQHL
PVNMOKLLNFGDGLPCTPAGIIELLNYYEIPRGPHAAIVGRSNIVKPLAALMMQKH
IYNTCTVTLIKQSENLEILKTADIIEALGALFLIKETMVAPHAVIDVGTTRVPADN
AKYITLLIDVDENNVATKCAAITVPGGVPMTVAMLSNTWRCYQNF

CpN_0336 380569 381591
yolJ
KDKMPPHHEHWHWHHARHAYHAYVMAKPKPFLVLLGLGIAHESQKTTTIEGEQMTI
FYKILVITHLIAKEKARLQDIDRCHKADIEYNNWNPYSELINRAPADVITLVEL
SEHLDVYVTLVYKLEGRDPITVPGIKTILWLLHLKQTLPPKQDVWQHYKDMQWQHLFQS
IKTKLIPNNHIVQIDLAIVVQYAVDQLEINTEFCNNYVWAKIKTKSGHHPGGRPW
IFKRAVITLIDILMAIATYNNHUKWVVKIYTHILDTRTKHLLELSSYPISQSVSVH
PKAYAJAIAVILMTTQYKIRAKWAEHIIIILTYIHDAQCS

CpN_0337 381575
smgB- Small Protein
IEEIPGNGQKRILILVLRPKNCLFLYFPLSGIMPEDLMAQKEIVSNRKLARNYEVITL
EAGIVLTGTIEKSLRSDHGNLGDAYVIVSKGEQWLLNASIAPYRFGNIYNHEERRKLL
LHRYELRLKLEKIAQKQMTLIPLOMLSRGVVVRLOCCRCRKAYDKRRTIIEEREKREV
AAMKRRHH

CpN_0338 382272 383375
CT119 hypothetical protein
FNMKFTLRNEIYKILKFTLITVNTTFLPLTHVLEIYHHEAVTATDITVATHTK
AKVYKGAISIPKRRFPQLKELTEANLEISSAGEMAJITGSSGCFLLSMEKEDFPM
PDQNALRSLPAEQKLTMLRTSFAVREESRYVLTGVLLAIVGVATVGTDOGRKLA
IDAEVNLKLSFSGEYIIPKAVEEIKMCSDEGAAIFLDQDKIAVECDMTLLITKLSG
EFPDFSPVISTESNVKLDHREELITLLKQVALFTNESSHVSXKSFELRGELTLTANCTK
GEGKVMANVYSSELLEIAFNPFLLDILKHSKDELVSIGISDSYNGIITDSASGLFVI
MFMRLHDD

CpN_0339 383405 384034
CT139 hypothetical protein
VTTLPMPKICSLKLNFRNHSDELSLAPKLNVAQKTNLLEALVYLSLGRSFRTOHLT
DTITFGSSHHFLETQEKDHLPAQLSIYTDKQKKICYNQLPIKLSQILGKVPVLFSS
KDRLLISGAPADRLFLNLLSQCDNHYTLCLSYHRAHQORNALLSKQSTSVASGMS
WSNTAPTYSNGFSVVRNFQIYKPNFGLTT

CpN_0340 383842 384156
(frame-shift with 0339)
PLYPLLIIVSSRSSEKCSLKQANLRGLWDEQLVKHGTYSIQRFCLCSQKLSDLKSEL
WSNLLKEQALKFKSSLIKNSIDSETAVAEFISLSLPDLE

CpN_0341 384160 384495
(frame-shift with 0340)
GSTSVPHREDFLLTMNMPVSQFSSEGGKSHLAILRLAECLYKSHHVSPLVCLDDI
HAGLIDNRVQQLDPAPTGLTITSTHMGELPKTSVLVSIENAVQVSEQII

CpN_0342 384619 385062
predicted OMP (leader (19) peptide)
HYKFKLTILFLAVGNPLFSETSVIQTLPSEIGGLKETSQKESVVCVHAFRSYTSKLP
IARVLEKHEVDFIWNVETKRTLEKHAHLNRLKKIAELKPGVPINFTVHSIGGVIVR
VALAHDCPEAKKEKLFWSLLRTQGLH

CpN_0343 384999 385595
(frame-shift with 0342?)
LPRRSQKRKAILMAPPNAGSTLARRYRCVKFQVFGKGLRQLLTTCYPTKMLNVGKLPS
SLDVLILSGNRHKSFLPFRPYENDGKVCITETKLTDPKHAIVHTSHTYIITNRKSLYL
MKFELKEGNTPIIEHVPEAALEQTVMEDKQKSNRLKYPNQDIYVHCFSRPNLYG
PKKWSLNQKNEINPEKLEK

CpN_0344 387432 385558
yaeJ-Metalloprotease
SSRYPTIIYFILAALGILVILHELGLHVAKAVGMVESFSIGPGALFKRIGGIEY
RIGCIPYGYVIRIGMERTKEKGEKIDSVYDIPQGFSSPKWRILVLVAGPLANILL
AVLAFSLIYNGGSKSNYSDCSKVGVWHPVLAQGLPGDEILTCNGKPVYGDOKMLTT
SLLEGLNLEIKRPGYLTVPSEKFAIDVEFPDKFGVPCSGASYLLYSNOVPLTNSPME
NSELRNDRFVMDGTLLFSMAQISQILNESYAFVVKVARNDKIFFSRQPRVLASVLYHTP
YLRLNLDITQYAGLKGKWSLYTLVYINSYVIEGELTAIDPESPLPOQERLQDGR
ILAIDGTVPVSGSDILRVQNRVSIIVQVMSQPELEEVNSRDADKRFIASYHSEDLLO
ILNHLGESHVPEVAGPYRLDPVQPRWDVYSSSELDKQLEVAKKIKNDKORYLELD
AEKQKPSGLSKDLKVRYNPSFVMSLNTIKESLITKALVTGHLSPQWLSGVPVIGOV
LHTGWSVGFSEVFWIGLISMLNALLNLPFVLDGGYILLCLWEIVKRRRLNKKIVERI
LVPTFTLLIIFIFLTFQDLFRFG

CpN_0345 388587 387436
CT345 hypothetical protein
LKVACLKHLAVLSTGSGIGRQTLVIRYRYPSEFKIISMASYGNLRLFFQOEEFAPLAA
AVYNEEYNEACORFPHMQFLQGLTQCLIMDTVTVAASSGIEALPALESMKKK
ALALANKEILVACAGELVSKTAKENGIKVLPIDSEHNALYQCLEGRTEIGIKLLITASG
PLLNKLEELSCVTQDVLNHPINWMSKSVTQYDSTLVNKGLEIIIEAYVLGLENVEILA
VILPQSLIHGMVEFLDGSVISIMNPDMFLPIQYALTAEPFASRDMGDFSKQTLF
PYDEERFPSIRLAQVLEKQSGSFFNAANEVLRVRFCEEISWCDILKRLTLMCEHK
VYACHSLEDILEVGEARLAQEI

CpN_0346 389690 388704
070-troD/ytgD-Integral Membrane Protein
KKGSLMALGSPSYGVSGFFQFSVFSRLFSQSLFTGSLYIDDIQIIVFLAISCSGAFAG
TFVLVRKMAHMANAVSHTVLGLVLCVCLTHQTLTSLGTLTAAAMATLGTFLIYFIR
NTFKPSEESTALVFSLLFLSLVLPMTKNAHIGTELVLGNADSLTKEDIIPVTIVIL
ANAVITIFAFRSLVCSFSDSVFASGLGIPILRVLYLIIQFSLACLVGAFKAVGVMLALF
LIIPLIAKVIKISRLMAWSVFSIGTAFAPASSRAILSAYDLGLSTSGISVVLTM
MYIVVPIFYFRGYFSKNFEKISEKSSQY

CpN_0347 391078 389678
069-troC/ytgC-Integral Membrane Protein
TFQTNPEALSRTIWIILMISCVFSDTIFLSSFLPTLICMTTALWGTILLISKQPLLS
ESLSHASYGLLVGALMAQVFSQACIFWIVLPGCAASVFGYGIIVFGLKVKCLHKDSA
LCFVLVWFVFAIGVILASVYKESPTLYNRINAYLYGOAATLGFLEATLAAIVFCASFL
WWWYRQIVTTFDKDAVTCGLKTVLYEALSIFISLVIVSGVRVSVGIVLSAMFVAPSL
GAROLSDRLSTILLSAFAFGISGALGSYISVAFTRAIIQGQAVPTLPTGVLVWICAG
LLAGLCLLSPKSGWIRVRRKHFPSKQDQELKLVFWHISHNRLNLSVDRFVCSYKY
QEVFGPKFPRWRVILEWRCYVKEODYVRLTKKRSALRLVRAHRLWESYLWESYLOF
SKESVHLEAEIEHVLTEELDITLTELINDPYDHPQIIPNNKKEV

CpN_0348 391815 391027
068-troB/ytgB-ABC transporter ATPase
EYMLLVVDETWSVHNLNLYNEAAVYIYIFGLGGLTALIGPNCAGKSTLLKASLS
ILIPSGQYIFPNKQKPKVRQIAYMPORACVIMDFPMTVLDALMXYKQKMGRIIS
DGRREAFHILERVGLCEVADROIQLYYXQYKAPALAKAMQKADLYLMDLEFIAIMAS
PYTGVLQELRQDQKTLVHHLDLHVRLFDHVVLLNRLICQGTQDEKINDTIFQY
KTCVLEELQTLKLSGKQFQZ

CpN_0349 392747 391790
047-troA/ytrA-Solute Protein Binding Family

ALKDPEGIPVIDTLDWVQNTAIDALNKGVPLESLIGAVLAREFQSWKEIREQA
ARNYPGTPLIFEMPHDPGVFTHALYASKIIGYAQGFMLDGEASKE/NMGLDLGEIA
LHWRCGCIQSAFLVIVHKGAANPENTSLIFQYFREGALRHAEMQWRRT/ATAIGAGLP
IPCLAAAITFYDGTPTASSMSLAGLRDYFGAHTYERNDRPRGEFYHTDWWHTKTTERV
K

CPn_0350 393169 393684
 720aa, hypothetical protein
 PPSMAAGRIHRIKRIKRIKALGKALPWPQINHHIDLCRYLFTNYYVYMHHTLKKMKGAAP
 EQLLKMLKTFLESETPTINPKPAPGRGSKRRDFINFTKTDIERVLELARQVGDKDLL
 ARFSPPKPLTSLKRELIIRISIRIGVSVELWNAVYEAVKAVSSPNLEVTSPFV

[illegible]

CpN_0351 393861 395432
 adt-ADP/ATP Translocase
 KIKVFRVNMHTKEEPFGKILRSFLWPINHTHELKKVLPNLMFFCITFNTVLPFLFALPT
 IYVAGPAGSGAAEIPYIKPWLVPQAIIPMLIYAKLSNLSKQALPYAVGTFPLFFIDFALPT
 VIYPLRDLHPTFADRLQAIPGLGLGJALIRNMTFAAFVVLAEHWSVNLSSIMWFG
 ANEITKTHEAKRFVLPFGIGANISLLASGRATVMSKLASRVSGCDPWGISTLRLLMNT
 IYSGVLMSVWINKVILTPDRPYNEPMQKGGKAKPKMMKSDSLVILARSPTYLIIA
 LQVLYAGICINLIVETKSQLQYKPMNDYSEPMFGNSFWTGVVLMVLFGVGNIRK
 FLWLTGALVTYPMVLLTGIVFVLPFIRNQAQSLVAMFGTGPLMLAVWVGAIQINLSKST
 KYALFSLTDEKAYIPLDQEQKGAALIDVVAARQKSGGKALIQGLLVGICISGAMPY
 LAVILLFTIAIWLVSATKINKLFLVKKALKEQVAEQVAGDSAPASS

CPN 0362 407843 407055
 fliA/rpsD-Sigma-28/WHIG Family
 LDKKKFVQKTOQTQNTIEMNFWTQETQIYRDSLIEFYLPVKSVVHRLISGMPSHVKTE
 DLYASGVQDLVRAVERYNPERSRFEGYVFLAKAIIDDLRKQDQVPRSVQKQANKISG
 AMDSLRSQCKGEQDLECELYNTSQGLSGWFSVPARALTVSLNEEYPSQSDGAGMAL
 ERIEPDERAETGYVQVKQFSLCLANATQEELEEKRVKMLYYEELVLKEIGKVLGV
 ESRVSTHISKALIKLRAALSFR

CpN_0352 395478 396380
No robust homolog present in Genebank/EMBL as of 11/7/98
NGVGFVLSHGFSTVYAFVKNQKIVITRHSGSCMKPCLSLTVPHVFLHGLDCEHRASCLKRI
RGLLGLLIGLVLLAVLSVGFVFLAAPSIVAGTGLAATLVIITLWALLAKSKVLPI
PNEQLQIYNRYPEKVFYVTKHSLTVKESVFLINCWKQTKDLPNHLHKAEGDILDK
SLDITLDFEPEEILLNCPWLHSHFDIKTESVAGEIGLNTKGVGLGLPFAHFKGYIT
IFSHYSTRLPILLISQSYFLYKASQNKQWDSPVSKTCEEIIFKELPHMIFKDVQVGS
QFLFLFSGHITWEGAQMTQLINPDNKMGQDFDKAGHCSMATFGFGLNMTIMEDPDVS
SNYEPTPNYFAHAKYKLVLEKESPMHPSALVQKICVITTHQNLLKRWQFVRNTSSQ
WTSLLPQVFTKYLEKIKLESSLPSSL

CPN_0363 407900 407943
filh-Flagellar Secretion Protein
EVNFVSGKQKQVGMIFVLSTLWIFLPLPQIDFLDGLCSFSLSLTVCWVFTLNSGN
SAIFNFFVFLCCLLRCLLRNLASTRWISQGTASSLIVLGSFSSLGLSWAATFACLLF
LQVIVNLSVSGSRIEASVRFPEALPAQMALDGLVSLVDSGRASYKAVKQKNALVEGDF
FSAMEGVFVFKGDAITSCILLNNVNVSTCLTYTSGYALEQMHFTGLDALIOBPFD
TSCAAETISPKIDKEESLLNLFYEYKQLRHFVRSVLLIFSLCCIIPSSPKFPIVLASL
LLAYLVKEPASEDSICERAFSYVEGACPKQESQVQYRAASEVDFDLGVLPLVLS
LIEREFLRVLVQGVNLYDEPVALPFLRNIAHALNAVQYVKYLESEGVFVAGED
IVPKIQLSSLLVTLRLLVRVLSLKFPLKFLIEAVNCGSDSLIEAEKRVSLSGWI
GRSLDKQKTEVLEITDHFVELISYSGKSNPMGVNIRRVDSLILERSVFKDFRAIVT
STCTFEMKMLDPHDFDPLHLSHDELPEKIPISFVGIVSDVLEVLV

CPn_0353 396893 397135
No robust homolog present in Genebank/EMBL as of 11/7/98
LRFNRNKKSLIFIKRIYRYSQSGKEQKGARPPFKKSITSSLVILLEAIFNENFSSIIQNN
FNKFNKNKNISINRIFVKFTI

CPr 0364 409954 410238
fer4-Ferredoxin IV
KENSMAKLVITSDEQOELEDNSEIAEPCESMGIPFACTEGVCGTCVIEVLEGRNLS
FETFEFYDFLGPEDSNERLACOCRIKGGCVKVT

CP0354 397062 398507
No robust homolog present in Genbank/EMBL as of 11/7/98
YKTKSIIKLIKIKTFLLLIGFLNLRNYQTIDEPRKMSNITSPVQIQRNSCNYFFELKNS
THTVFTSAILLGCALIFLCVAFSPVYLSLGAQLGLLIALIGVLIGIKKTPMISKE
QVQPEQLVNRIRAHYKPFVDSFSEAKPNKDLISITIDLLNQLHSGVSTNMYNSELQ
KQKTFEGTARLKNQEVTSALRKLESAASRFLPKLPIKQVFFFWLGEFISAGSK
VELHVRKVIIGSLSELDYSIKPEMLPTWYLLPDRFSTNSINLWTLVLARVLTRDVF
QHLYKAAALNGENLNLSDNLTQKQLFARKYAAQYOSYKHLSPQSEDFENKLLIKFKH
RYSKMKSILKTVPDLWNLCCLTLDHTRGPQDMEFASIGLTLYTQGLIHCSSEAFSL
LTLELDQFKTIRROSTNIAMLENLANTNSTFRSLPPTIVHPLKRVSPQSEDESSL
IG 524aa

Fm_0365 410498 411544
 No robust homolog present in Genebank/EMBL as of 11/7/98
 FGTSTLNGSLMATISPTSLTVDPVSLDTKKKSNFDDIKQSRILLITAFVLTITGTL
 IGLNIPVYVLTISFSLVSNFLVRAITLLKPRACGKHKKPKRPNSTNLQYS
 ISTAINRSKNEKNEHQDKLNPAPSSALLDNPYEVLWAKHSLFSLVSLPLGNGPHEHLL
 SASGLNDKLTLLIETSPNQATISSVYVDTSPSKSLNLNEAIQTRTEINTELPAGSDGRSLY
 RSPDFGRVFLPVIQPTPEATVYVYVYVLTQATINTNQIIQIPLSLRHHYSREL
 RSPSPMOQSIAMTAVVYMAELFPEVPLTIACTVRSKAPQESIEDLS

Cnfr355 399955 398591
 No robust homolog present in Genebank/EMBL as of 11/7/98
 IDPGLYHIITAFNRSIKELAMSLITVPHALFMNCECHSTPFLSSRTIVRIASLFLD
 IGAALGCLAPPVSYSGVSLMTAFIVLITLVILALFIGEKKLPPTPTIRIPDAPHFVC
 EAGLSSISAFVREQVTLAEFRQFSTALLCNCSPEEKIKQLPSLRSKYSFGSLRGAD
 LRNNWPIFDLLSTCPLYWVLQIISAGDPQVREKQVVPRECYGYWGLPGYSITAKAT
 IFKETHHILQQLTFLDKLLLNKALQEKWDTDEKVAIVERITYITITARGTLTEAGGLT
 KETISKELLLSHGYFSFDQLITQLPRADWLKCFVDNSTAYNLQCLAGLVSQNL
 LDDESIIDFVNLGLVYLDQLEKAAVFASSDEPKKELGKFLRLHLSSVKRLESVLRRQL
 RLHIEHGNAARLVYDVTGAEIRHRSKTSIFFDK

CPn_0366 411976 412440
No robust homolog present in Genbank/EMBL as of 11/7/98
MGYLEPVSATDVLFPSPAPLINSANTONKQKLIELKGKQQAESSPRTITSVILEVLLVIGC
CLTVLSLLAIRPALQFTLETHGPAAI AVLAVSGTILLVAVIILFCFLAAVPAAYAKCKTYKY
KTVLVVDAWSHSHOOTPTLGTIFSGIIVYAESAOL

CPN=0356 400465 400109
No robust homolog present in Genebank/EMBL as of 11/7/98
KQVFLQYMNESGGWDLDFQSGEGFQLSRVLGLLHSSWALYEAKQFYLPVSLTWE
FVDFVDFLISKPTKHGKVDKCNVEKHFQFRFROYLGLSLDNORFENTFLNYPKYHLDR

[illegible]

CPn_0357 401341 400469
No robust homolog present in Genebank/EMBL as of 11/7/98
YSSNHGASMVNIQPVYNTQVNIYQATQFSVCQPALSLITVSVAVALAIALVCSQSL
SIELGTALVLVLSLIFASAMFVIYKMRQEKLLIPKKIMLEIQHYPSIVQVFDIQVE
SYIEIHLISILNKNTVFDKAPVILQEKLLQFGIEKFQDVHPSKLNPFTEILLQCLPLW
LGRVLVPMVSDVTPCTYQYVSCGPLGLYENAPSLFFERRSLLLKKISFGFALLDGLKK
MTSSGSSVQIWRNLNLTFRYADQAEFVDFAIFNADYQFDSLLHLISFKLS

CPn_0368 413766 414107
No robust homolog present in GeneBank/EMBL as of 11/7/98
TLAKDYLWVNAQHQSFIETGRINDTPNGEAHFLAQLLGPKYEGELKAHPKLSNVIKK
YVNCDFEALNNOATVVOVPLISSIYSPGGKLLEFPVNOTKPNSSAYKLYHIRT

CPn_0358 401757 401578
No robust homolog present in Genbank/EMBL as of 11/7/98
EEVLVSCMKL IPTDLSIERETDSKRDKKIETLYICSSKVL AGHEESHLDKHKI HESIGV

CPN_0369 414345 415562
C0508 hypothetical protein_2
NIMTDSNPLPSTYASLYRTAKHVSPIRLPLNRDTRIEKILKIVTLTLALACALGFSA
AGILAFMIPSAVVVLTALTAAYVSLSKPKPKYLEIQLPQEPSSQSLSPSPQPPEQD
LPLQIDLPDLPSPSPLEVSADLTTPPEELTAITPTVGYEALQNNQLSDLPSPSPST
TPTPQPCFVWLKDKSLKIFISGSDIAPVRIKTQGRVMINVAENISREGGQNTKALS
LATSLCQWNASRLPRAHSRSGLOQDGCERSAKWSNDSHTSDNHVPGKAHFLAQLLQGEA
AKCNMDPKQAFESVKAHFNLFQAEITGDVDTQLPLQCNLFPASRLNLNGKTRAEWIE
AIIALITSLQDGEWDDNQFEAKIIITGDKDPIIPRFDLTP

```

CN_0359          401994      403817
ClnA-GTPase
ITPOLYILKEYKNIENRFSIIAHIDHGKSTIADRLLESTSTVEEREMRQQLDSMDLRE
QSTIKAHVPNTMYLYGEEQQLNLDTPGHVFSYEVSRSLSCAGEALLVDAAGQOVA
RGSLT(LALDERLEI)IPLVNLIDPAADPVRIAQQLIDTGLDITNITACSAQTGGQIP
ALILKAIIDGLVPPPKAPAEETKALVDFSDHYDPYGMVYVRIISGEILKGRDIRTFMAAG
SSEFVGLDGLFPLKATIEGSLRCQGVGFITANLKKVKDVGIDGTVTKTKHAKPTPLGFG
KEINPEYFAGFVPIIDSSDPTDLKALQVGLNDSALTIEQSSHSYCFRCGFGFLHHL
IEIPEPIAREFDILITATAPYIKYVLLKNGKVLNIDNSPYQDAIIEHVEEPPVHNI
ITPOLYELNIMNLCLKRKE(CVKTEMLDQHRLLVLALEPLNEVTSFNDNKLKSYTKGYS
FOYRLSYDHYKGS)[LKLVEINLEPDAFSGLVHRDFAESGRGSICEKLVYIPOQLFPI
IQAAITPVKYIARETIRALSKNVTAKCYGGLDTRKRKLWEKQKQKKRMKEFGKYSIPNTA
FLEKTLQD

```

CPn_0370 415755 416912
 C0508 hypothetical protein_3
 KRPFKFLVYFKLSFMSTFEPNLWNVNLMTLISSESMPTQLASHKLKGLDLVAFILITIGI
 AVSSGATLAIIGLFLITLALVAFVLAFLSILGIFLLRPEKSPISVYHTPTPIIKDTDLPPV
 PLLALTVPVTEAALEELPPFLSPRTHTLQENWDRIPLDQANTDMFFIAADNQTGYAWHL
 KSNLNLTLSTGLGIEKPRYKATQIVMIVNAATPMANNVNQTSLAKALTSVRVENSEK
 PSDLRSGQLQGLGECRSAWENLKMTKAGKAGLQFQLGQLGKASADYNNPINDAFTF
 CRYATVNLCLNEAKRRKTTVQLPLGLSHFPCPKDEETSLRIQLDGVKLALIDALQTL
 GSEAEQNGPQWVILTLTLARPLTTP

C1n_0359 405364 403922
 VAL160 hypothetical protein
 VALTLLPGLGLAVMKKNLVLMDHGGFSVGVNPTFKTRDFLKEYPNHRELVGFESE
 DEPNLEPERHKLTMLDIAQKPKVDQTHALLPFLLEGGVTLDFQSNYPKDSERPKKELOEK
 GLLPFLTTHAKGGEIARIKPIETMRKNPNPAWLVLPVIFQSLAAKVGQRPCCSWVGTQAG
 EYKAVAKVETFGVDGLDFAVYKILDFELKISATPAVATLKLEWNTLELSYLRIGAEVL

CPn_0371 417141 417501
No robust homolog present in Genbank/EMBL as of 11/7/98
KTMVPSSAPLPTSHRPSSGNLALMEPEKKAFAPIHQDPTTKTKILVKILVAIVLVIEVLG
LTAIAFFIRGTPPLCLILCANILTTVAIVLLVYIKLALVHKTEKTPARQQIKRKLSEKSI
S

CPri_0372 418651 418661
No robust homology present in Genbank/EMBL as of 11/7/98
NYVACHRNIGNHSSPVVVTGSGAGPVQPTPLGEPERLGTGGRCKIAFAAATALLLN
TEVGVIGIAIMIEVATSVGAYETPLGELLELLLELLA(ML)GMYKTHPSQNTPLGN

GIFYAFSPNLEVTZILSME [REDACTED] GYNADLWZKFCF

CPD 0448 504876 503628

[illegible]

CPn_0449 / 507231 505330

pmp-10-MMP-10-(Frame-shift with 0451)
 eayTgRRGGGGSNNIVGTTAGNGAAGSILAEGCSLSAEAGDITFNAGNAIVATTPO
 TTKRNSIDKSTLTLNRAISGHSIFFFDPIITAAZADTDLTNLKNADGANSDDSGS
 IVFSQIGLSEDAKJADNLTSLKQPVLTAGNVLKRGVTLTDKGFOTAGSSVINDAG
 TLTAKAEVTEVLTLSLPVDSLGGBKVVIAASAAQSVNGLSGPILLDDNONGNAENHDL
 GKTDQFSFVLTALGATLTDPVAVPTATTHYGVQVQALTDSTPIKTKATLA
 WMTGTYLPNPRGOLPVLPNSLWGSFSDQAQIGVRSALTCLSDVRGFAAGVAVNLFKD
 KKGKRIKRRHSGSGAIGGAAGTCSENLISAFQCLGSGDKDFLVAKNHTDTYAGAFYI
 HLTGSGSFLCYLLDLPGWSGHKPVLEGLQYASHVSNLTKTYSYAPKVGSGMGNNAF
 MMLGASGSHYPLHCFDTAYPIKLNLTIRQDSEKKGTSFDDSNLFLNLSPIGV
 KEKPFKDFSDYDLSLTVPPDLIRNDPKCTALTIVSGASWETAYANLRAQLQVRAGSH
 YAFSPMDFVLGQFVEVRGSGSRLYNLDLGGKFOE

CPn 0450 508121 507180

pm_102 Polymorphic Outer Membrane Protein
 SCKMFKQISVWLVLSTCLACFTSCSTVFATAENIGPDSFDGSTNTGTYTPKNTTGTIDY
 TLTGDTLTQLNGDSAAITKGCFSDDTESLSFAGKGYLSLFLNTKSSAAGALSVTTDNL
 SLTQVLSLTFLAAPSVTITPKGAGAVKCGDITFDNNNTLFLPDQYCEENGAISTNTKL
 SLKNTGTSISFNGKSSATGGKGAACATGTVDTINNPTPLSNIIAEAGAISTNGT
 CTITNTSLVSPNSVNTATAGNGGALSGDADVTISGNQSVTFSGNQAVANGGAIYAKKLT
 LASGGGVSPFLTI

CEB 0451 508158 511058

10P-10-MP-10 (Frame-shift with 0451)
 TPTQVRKIKLDCSCFVFNLFYLCFYIDANSSLNQKSIKTKTSIPWLVSSVLAFSCHLO
 SLANEELLSPDDSFNGNIDSGTFTPKTSATTYSLTGCVDFYFEPGKGTPLSDSCFKQTND
 TFLNGHNSLTFGFDIADGTHAGGAATANKNLTFTSGFSLSSDFSSPTTVPYTGQGLSS
 AGGVNLENIRLKVAGNFSTADGGAIKGASFLTGTSGDALFNSSNSSTKGGAIAITTAGA
 IANNTGVVRFSLNSTASTSOGAIDDEGTSILSNKKFLYFEGNAKKTGGAICNTKASGSP
 LITISNNKTLIFASNVAZTSGOAIHAKKLASGGFTFLRNNVSASTPKGAISITDASG
 TSLASLAEATNGTIVFVRTLLTGTSTDPKRNAINIGSNGKTELRAAKNHTIFYDPITSE
 TSSVLDKINNGSAGALNPYQGITLFSGETLTADELKVAIDLKSTQPVLSLQSGKLLLO
 KGVLTLESTFSFQAGLSLGHMDSGTLSTTAGSITITNLNIDVSLGLQPVSLTAKGASN
 IYVTSVKLNLITLIEGNTYSFMSHFDQLSLKLKITVDADVDTNIDSSLIPVPAEDPNSF
 YGFGQGMNVNMTDTATNTKEATATWTKTGTFVSPERKSLVNTLGVGFTDIRSLQQLV
 IIGATGEMHKGQFWSMTNLFHKTGDNENRGKTRFSGGVYIGGSAATPKDDFTLFAFCH
 FARDKDCFTAHNNSRTYGTGLTFPKHSTLQFONYLRLGRAKFSESAIEKKDFPREIPLAD
 YVYVFSHSDNRMTDGYLSLPESEGGSWNECIIAGGIGLDPVSTYNPHPLFTFVDPIMQVKE
 YVYVSQNSDFPESSSDHRTSGILGNLLSLIPVCAKFGQVIGDSDITLDSGFFVSDYRNN
 QSTATLVMSPDSWKIRGNNLSRQAFLLRGSNNYVSNINCELFGHYAMELGRSSRYNVYD
 GTKTRF

En 0452 511304 512860

mp_12-Polymorphin Outer Membrane Protein (truncated)
 MEETMTITLIRNFLTSAFLALPAAQVYVYVHSDGYGAINMKSLEKPTCYPEGTSYI
 LLDVRIISNVKHQDEQDAGVFIRNSGNLPMFGNRCNMTFFHNMLTGEGFAAASNRVGGDTLL
 SNFYSLAFTSAFLPGOGGATVYSLGSVINSESEVTCGNYSSWSGAAITPYVLGSL
 PFRVSNLSGNRYVLPKDNVSSQCYGGAITTHNLITTRGSPSCFNHAYHSDVNSGGAIA
 IPGGSISSVYSGKGLIFKGNATQSGDNTNHSILHQSGAQPNKLRAVESGVGYPDPTSH
 ESHKRIITDLVAINPEKGETYVETFSFGLGLDDHVECAENLTSTILQDVTLAGGTLISLD
 VITQLGHSFQKEASTSTMTSPGCTILKSGDARVGNHILIEDTNPVPRVIRAEKDKALV
 LEKLKVAFAEYVPSLDVDPKFAETIPLLELGLSGDSLLLGTTTLERTQVTTENDAVR
 LSSLSWEEYVPSYLDKDRITPTKTVTLTNPETSTP

Pn_0453 513156 516152

MP13 - Polymorphic Outer Membrane Protein

CVLLYLFYFYSLSLCRIIWHFLVYOMKTSIRKFLSTLTAPCFASTFASTVEVIMPSNF
GSSGKIFPYPTTLDSPRGTLCIFSGDLYIANLDNAISRTSSSCFSNRAGALQILGKGGVF
TLIRISSADGAIASSVITQNPCLPFGSFGSGFMFDNCELSITDSASNVIPHASATY
TFLMPTNNDISILFYQNSRAGSGAIAIRGTSITENTKKSLLFNGNGISNGALGTGSAAL
NLINNSAPVFTS*FNTAGIYOGAAILTGCSMLTSGNLSGVLVFNSSRSGAIVYANGVNT
GMSNLSDLTQMTASQNSLPAPTPPTPPAVPTLVGLYGGAICTPACTPPPTGVSLTIS
TGLNNSVNTFLENTASCGALGYGLNISIDSNKTLFIGNTAGTGGAIIAPESGSLSLASNOG
ILFNKNSLITSGTFTNSIHFGKDAICFATLGTAGTGLYFYGPDIYDLOSLDASAASAAATV
VPKASADGAYSGT*VLTSGTDLTATEAATPANATST*LNQKLELGGTGLLRNGATLVNHN
QDEKSVYIMDAIT*FATTEGNTNMGATILANKLILNKLSDGTGAIVYVQSTNGATL
TGLTGLVKLNSQD*DNHGMFKNKDLPQVPILELKATSNITVTTDFSLGAVVYQOQSPVGYQ
WEETIDTTHTITGMKKTKGLYLPHERLAPLINSLWANNILDRAYVCSAADAAGEDVP
QKLSITITNFFAHNNTGDARSYRHMGQGLINTYTRITPDAALSGLGQLFTSKSDYL
HGHGNSNVYFATV*NTIKLPGSGRSFGSGTSVYRSGNEKVYVSTIKLPGKRSWNS
VGLTEGELNLPIT*LSRRIANLQKIIIPVKAEEVAYATHGGIIGYDTEPIGRILFHGHLNVA
VSGVFPKGNKSHNR*DFVYITVAYAPDVYRHNPQDLTLPINGATWTSIGNNLTSTLLV
LCSGT*SVNDVLEI:FGHGCGDIRRTSROYTLIDGSKLFF

rn_0454 516179 519115

_P14_Polymorphic Outer Membrane Protein
 PLFLKSPFSCCLALGASCAFAETKGLGIIIVPPIITQGEILLIDFQVFNPLGAGF
 FQIHNSNLGLKLLFTFTQCAQITNSIIALLSAETLTFKNPSINFTNQCTGL
 LTYKGLVDFPSIKDLITFTTPGVSAGVITTSATPIATITVTVTGAGALVITL
 FQPFEGNLANFSGALGSPFTAAVKFTNNITATMSFHSINFTSSQGVVIVKSL
 LFTFAKCNHISKVDTSTFALGAGGAGATIPGTGTEKLNKNGKTFPTVITNDAG
 VYAEPTNVKNAALLDNTAPNCAIAKAVNLIGRPIEPRNRNKAQAIETP
 FQPAKQITLTLAGRPIAFQGNILKPTQIRNATLYEAGIEVLSAGVSRIVFY
 ITGILGTLTSSNDITINAGIAEGLVPTPLTSLTELLPHTMTITLTKVLAGGE
 TIAAVITLAFATYKATITKLTGTLGATGTAAPVDFTLKLAFADFTPLPD
 GAGTAVITKPTLTALATDEHATITATLQGVIAETLAVERTAVPKPKFQGE
 TPGYGVYKWNSTSRNLLIPAFVITSPGSPGANTLYAVNINILVETVITL
 FETENILWISFQAGFADITDIALITLGLSLITAKALGAYVHTITPHTETG

CPN_0454
No robust homolog plot in Genbank/EMBL as of 11/7/98
SLETRGRTEICLGLLFFDIDGLKFLQLGEGTALNLRIFAPLRNRTTESRRARQPD
HRIAIYIVGLDSESGKILERLISYMSGLYSESGMYLRFFMGKNQNGAVLSKLVENLH
IRCGFFSEDAPVEPEFDLSIIVHTDRSCP LTKKRSWSWELQTVLPESYIQSEFMLS
RFRMLS

[illegible]

CPn_0466 533718 536537
pmp-15-Polymorphic Outer Membrane Protein
TSMRFCEFGMLLPFTFFVLANEGLQLPETYITLSPYEQAAPQVGFTFHQNQDLAIVGNHN
DTFLDYKYRSNGKQALTKCMLITSENIGNVFKEVNCVNSGGAIYAQNGKITISKNQYAF
FTNLVSDNPATAGSLGLALPAICNITNLNGQOTFVDNLALAGKQALYETTNLSIKDN
KGPILIKONRALNSDSIQGLSYGNSNLITENSGNAISITSSNSGSGQIFSTQTLITSSN
KCLIEISENSFAFNYYGNSFNPGGGGLTTTFCLIANREGVLNNNQSGSGNGGAIHAKS
I IKENGVFPLNATAGGALLNLASGSGNSFLISADNGDI FNNMTAKSHALPNPYRN
A HSTPNMLQIGARFVYRFLDYPIHEHLPSSFPIITFLNETGHTGVLFSGEHWQNFPT
DEMNFFSVYLNRSELQAGDVGAGLCAKYFQRGOTLLGGQAVITTAGTITPSTST
TPTVGSTITLNLHILDLPSLTSFQAQPKIWIYPTPTSTEDSNPTITSGTLTLRNS
NNEEDYDSDLDSHLEKVPLLYIVDYAAQYNSQDLSLSTNSGHHYGGYQINSTWYET
TITNPTSLGANTGKHLLYANMSPGLQRPHPERGEITNALWQSAITACLHSLSWYET
DEEKGHAASIQGLLVHQDKNGKFGKFSHMTQYSATTETSQS PNFLSGFAQFSSK
KEHESONSSTSYHFGMCIENTLFEKWLRLSVSLMPTSEHTHTMYQGLLBNQSGSSK
HNHTLAGALSGYLPDPHGESQTYPTFFILAIQRNALAFQSGDHAREFLPHRLPTDVS
LPGVARSANLHHRVPLWLTEIYSRSTLYRQDPELHNSKMLISQGTWTQTQPTVYNALG
IKVNTMOVPKVLTLSDYSADISSTLSHLYLVASMRFL

CPn_0467/ 536528 539434
pmp_16-polymorphic Outer Membrane Protein
NEILITISQNRKIKEPLVSTPKPKFLFLYLGNQACMFGMTPAVYSLTQDLSKFLALERDE
EFTKSTPLDLSLITGFSPTVFNHNRSSQDILFNYSKIDINILLTTSAGGAVSCN
NFLIWDHAFHSKSLNLAIGTGGALACQCACTITKNRGLIFPFRNRLNNASTGGETRGG
ALACIAAGDPTISQNGQPTVFVNVSNNVAGIQTSTVAIKNNSGVSFNNMTLSGINSNGS
RSEDTITISDNRPIFYKFNKNNVAGIQTSTVAIKNNSGVSFNNMTLSGINSNGS
GGATVYTNLSISDDNPGTILFNNNYKIRDGGAIQCTQLTILKNISGVHYFTNMQNGWGLML
LDQSTLSLFAEAGNIAPQNFVEVLTFTTPGRYNAIKHTPPNLSQLGAKNGYTTAFDPIEQH
HYFTNPLFNPNNNAHQGTILFSSAYIPASDYENFISNNKSTGKRLNGVLSIDRAGWQ
EYKPTQKGKILGLKHAASITANTSEPTSVSGVSGYINLNLINLPSLIAGKAKPTLWIR
FLOQSAAGPTEDDNNPTITLGGSLTLLEENEDRPDYSILDSEPLQNHLLSLSDTPARHNT
DNFHPESLNATHEYGVQGSWPYWEVTTITNNASLETANTIRALYANVMPTLOYKVNP
YQGDLTALTPLWQSHFTSLSLRYNRGTGDSIERPFLQIGADGLFVHQNSIIPAGQFR
ISTGYSLQASSETSLHQIKLSGFAQPTFRTRKEIGSSNNVNNVTSLSIVELPWFQEA
ATSVLAYGVGDHLLHSLSQKQGEAGTCTYSLTLAAIKGSPFVQKQSLYHLSPFVQA
AIRSHQTAETEEDGNRPKRVSPQFPYNTLPLGICQKQWSKFCFFMTLELSYQNPVLYQ
QNPQIGVTLTLAGSGSDWIDIGHNYRNALGYKVNHTQTLFRSLDLFLDYQGSVSSSTTH
LDQSTLKF

CpN_0468 539608 540432
pmp_17-Polymorphic Outer Membrane Protein
IYKLLDNKLMIFYDKLVFHKVFMFMRPICLSILSTALCCSLSGNEVPNLASCSMSRKDI
SAPHTSPSFLRALVPELVVSPFSRPNLSGLVGDITQDITITGNSIVSDYNYHEDGG
LACNKLFIYSENGKNLSFERNSSSSGALVSYFCTSKNQNYISINAASSLATTTTSG
FGGAIHRAVSYITNNLGEQDFLNDWSKNRGGAIVGVVSLSDITNLGPIVIKKQNTLEDSS
FGGQIFCRALNIEARNYQNTINDNAGSGGVVYFLP

Cn_0469 540399 541460
pmp-17-Polymorphic Outer Membrane Protein (Frame-shift with
0469)
CFTFRGGIFSLGVIITSSNKEI EISNSASSINTASGLKYPGGGIMCTSLVIENPNPKG
LIFNNKALDSGGAITHRSFIFQNGPTAFINNASTSGALINLGGISTPNQFLSADY
GDILFNNTITSSSPQGYRNALYAAGINKLKGARQGYKILFYDPIDHQDTTDPVFN
YEPHLLGLTVLFGINDVSNATNPLNLFSGKSNSSRLERGVLAIEDRAAISCKTLSTQGI
LRLQNALNITKPGGSSINFNAIINLPLSLQSEASAKPNFYPTLTGTSYEDTSSIT
LSPGTLFNDENINPYDSLDELSPRKDIPPLPRCCDKKRYFESHCSRHSLER

CPn_0470 541357 542532
pmp_17-Polymorphic Outer Membrane Protein (Frame-shift with
0470)
ISLNLERISPLLYLLDVTAKKIDTSNLIVEAMNLDHEYGQGIWSPYWMETTTTTSSTVP

SLQGLGLLINOHNREGKGRFNHTTGYAATTSAKTAARHSFSLGFAQMF SKTRERQSPST
TSSHHNYFAGLRFDLSLLRDFDI STGLSLGYSYGDHMLCHYTFILKGSSKAFNNHTLVAS
LDCTFLPARITRTLELQPFISAIALRCSQASFOETGDHIRKHFPHKPLTDLSSPIGRSE
WKTSHHIMPLWTTIEISYVPTLYRKNPEMFTTLLISNGTWTQTATPVSVVAAKIKNTSQ
LFSRVTLSDLSYASVSSSTVGYQLKAEHSCTF

CPN_0471 542561 545401
pmp_18-Polymorphic Outer Membrane Protein
TVNNRSLKSSSFVGLTGKTTILLNATPLSDYFDNQANQLTTLFPLIDLTLMNTPYS
HRATLFGVRDITDITVDLHONSISWSENFSDGGALSSKNTNKNQKILNLSNFAI
KRAGAMVYVNGFDLSENHGISISGNNLPSNAGFADCTCTGGAVLSTKNTISKNOGT
FINNNKSSGCAICAAIINIKDNTGVPFLFNNAAGDTAGGALLPANACIRENNSPQIYFLN
NOSLGGCAIRVHCIEILTKNTGCVFNFLNFAEADISSHSSGCAIYCSISKDNQPIA
AFNNNTAARDGQACITQSTLTIQDQGVFNNTNGTWGAILMRDQACTIADAGDGIIFY
NNHFKFTDYSNIHVSNTNRMSLTGVASQIGATVATDILQRYTQNSIQKFNPNPHGL
TLFSTSYITPDTSRKDDIFSHRNILGLYNLTALLEDRAEWAKYKFDQGGTALGRSA
VFSTTDREDSKQVSTNININLNLINLGLVNRVAPKLIRPTGGAGVQYGGDNNPILNL
SGHMLLDGNNILPDYSDADLACHIAEVLPLYLLDTAKHILINTNFIYFELNLTAFYGYG
WATVPWITETITPTDPTDADNTLIRQLRPTWPTQYVFNENKRGIALCAFWQYSHNLE
VLYRQYQQQQAITATGATGATPLVHQQNDNAGDFTMEATQYGLATITNTANIRPGVN
FDGLRPNLYLHESNPNVASHITZALQINNIHQERPESSATLASYVYINHIKAYQYK
ITQYKCYSTPLAALQYSLQWLRISHTITFQALAVRNSNAPUEGSGDQKRRKSVIK
KGLNLTNYPLITQISAWCKEPLPFWYHILLAYQVLYVQHPPEINLNSLESGYHMLLQZTT
LAHMLAFGRNQILFPKILVFLDPLVQKSSSTTTTTLVDTGTPFKF

CFM_0472 545514 545581

CPN_0546
r121-L21 Ribosomal protein L21
LSKQRLTSLIERFIRKKLMPEYAVIQTSGSKQYQVRGGDVLDVLLGCFAGDKEVIFQDVL
FVFDGTASLGSPTIANAQVKAELYSHVKGKVVAYKYKKRKNYHRKKHGRKYLRVKIR
ETLI

[illegible]

CPn-0548 33234 632191
cysJ-Sulfite Reductase
KMYLQKCEKAGQVPTLAEILLSSDKSDINDSPYRMVFNDSNDTTISYKVGDALGVLPENS
KVEYSEHSEKAGQVPTLAEILLSSDKSDINDSPYRMVFNDSNDTTISYKVGDALGVLPENS
KVEYSEHSEKAGQVPTLAEILLSSDKSDINDSPYRMVFNDSNDTTISYKVGDALGVLPENS
CSFLLCSLQVNDVSIYFVQPTKHTLITTEGKPLVMIGATGVIAPYKFAELFRNLND
PQNNLLPFGFERKNGFIYRFYNHMAEEEGKLKLFASFRSDRQKQVQDQLRIQKDEVYR
KAYFEGGFEVCGKVLGIEVHXALEELIGKDTLASLRKHRYVVVYV

CPn_0549 633662 633255
rs10-S10 Ribosomal Protein
PDQVQHQPWNGHSLRFLKKFKRLLRSKCGCMQKQKQKIRIRLKGFDQGGQLDRSTADIVE
TAKRTGARVVGPIPLPTKREVVTVLRSPhVDKKSREQFEIRTHKRLVTDILPTGKTI DAL
KMLALPAGVDIKIKAA

Cp_n_0550/ 6335688 633580
fusa-Elongation Factor G
LYNGNKGFMNSQEPDLSAIRNIGIMAHIDAGKTTTTERILFYAGRTHKIGEVHGGGTA
MDHAGEQERGITITSAATTVPWLKAGNIITDPGHVDFTIEVERSLRVLGDGAVHFDVMS
GVSEPTVWRQADYKGVPIRAIVNKHDRMGADYFAVESMKCKLGANAFVPHVCIGSES
QYVGVLDLISQKALFYLDLDTLKGAKWEKIESDLKFCARLEAHLNLEELATIDESNAEP
MKVLEPDSSTIDEIHDVGKGVKGINENFVLCGTAFKNGVQQLNVAIVKVLPSPLDRG
NIRGINKLTQDIEISLEPRRDGLAALAFKIMTDPYVGRITFIRIYSGTKLGGKSAIINST
DKKIERISRLLEMHANRDRDEFTDCTAGCVLGRSVGTGLDCDDNQEIVLERIEFPST
VDMAIEPKSGDKREXLAQALSSLSEEDPTFVRVSTNEETGTIISGMGLGDLTLDRMI
KFKVGNAGNKKQGVSYKETIYVSGNSETKYKQSGGRGQYAHVCLIEPNEGKGNVEVS
RETVGVVPKEYIYPAVIGKIEEGLNLTVLGAGVGLDSSVSVFVGSYHVEDSSMAFKICGS
MAVMDCAKRAKVPVILEPMKVAITDEPHLDGLTGLDNLRGKAGLIGQSSRGMQVNAEV
PLSEMFQYTSLSRLTSGRATSTMEPAFAVKPQVQIQEEIVKGI

CPn_0551 636174 635698
rs7-57 Ribosomal Protein
MYMSRRHSAEKRDIPGDPPIYGSDVILEKFKINKVMHGGKSKVARKIVYSALERFGKKLNLEN
VLEGFGEALENAKPILEVRSRRVGATGYQVPEVASERRNCLAMQWIKHARSKPGKSMK
VGLATELIDCFNKQAGATIKKREDTHRMAEANKAFAYHKW

Qn_0552 636698 636219
rsl2-S12 Ribosomal Protein
IQAGVVPSSSENKPLPTKRALLYISMLVVVRLKREYMPNTINQLIRKRRKSSLARKKSPA
LQKCPQKRGVCLQVKTKTPPKPNSALRKVAVWRLSNGQEVIAIYIGGEGHNLQEHSTVLQ
GGRVKDLPGVRYHIVRGTLDCAAVKNRKQSSRYGAKRPK

CPn_0553 637753 636812
No robust homolog present in Genebank/EMBL as of 11/7/98
CGHVVLRFLRILFICLGRVFLRASESPWSTCTLTGLGPIITDITNEFDVAQCG
LQIGTISNNNAIKIEFLIYKCPKEASIFSKRTEPLNLSQSHSLDGLICMNGEYTA
EGMKANEGKALPKQKDLRLVLRCPNPDTLIYSEKAEKGIETNCTLCNGQYTLDDGL
LYGDS IEKFLKPKRNNKHITLVLDSCQVVTTFELNLSLVNLYVFLSEDSAKILAI
IPDLQAATQLLSHTVPLLFITYTNDIHIIEQKSESSFYTNQDLTEPILGFLFGVINRGM
EYCFNCAQSLGET

CPn_0554 637806 638141
CT440 cytothetical protein
VFSLYLLCIIIVYVRFMYEGKSRMASPTPGQLHLQOKVESKAYDYSRLAMIATALLFFI
VALILSGLSLLPOVFLPFGSAYFIIGSFALFIALGILLINCVCDLKQYLTSS

CPrn_0555 638298 640241
tsp-Tail-Specific Protease
MFVMKMLVRLCVLLSLLPNVLFSSDLLREEGIKKMMDKLIEYHVDAQEVSFTLLSRSL
SYQSGDFPKHSLYSNQEAVYISQSPKTRKRLLNKYAGNFATYRNINLHESILRAROR
RNEWYKDPKMLVEASYSQIKSPQMWKSLDEYVKORQALLSLYSLHLGASSRYSYG
KEQELAAALCRLEINENHYVLDINDHGVAAMDREEAYQPHIRVVKALHSLDAHTAYFSK
DEALAMRLQEKMGCKIGVVLKEDIQGVVREIIPGGAAPKSGDLQGDIIYRVGDKGT
HLSGFRGLDCLRGHGGHVDLIRHGESDHTLALRREKILLEDRRVSVSEYPYGGDVGIG
VTLHSGFYEGENQVSSGDLRRALIQKKEKNLLGLVDIRNTQGLFSQAIKGLMFTNG
VVWSYRDAEDTMCKYVTRSPKPFYGLDPLAILVSSKSSAAESAIVATQLQGVGVVLDQEQ
TYGKGTIQHGTITCGASQDQCFKVTYQGVYSPSKSTQLOGVSKSDILPSLYADRLRGR
FLEHLPADCCDNVLHPTDLDTQTRFWFYQYILPNLQKQETLWRMLPQLTKNSQRL
SENSLEPLDQITKSKSEDTLQSGYNDNLDQESINLKDMLLQQCRK

CPN_0556 640921 640325
 CpnA-15kDa Cysteine-Rich Protein
 ENGMSSNLHPVGOTGTGAAPESVLNIVEEIAAGSGVTACGLQAITSGFMGNLLIGWAKT
 KFTQPIPESKLFGQACQITLLVLGILLVVAGLACMFIHSQLGANAFWLIIPAAIGLKL
 LLVSLGCPDEACTSEKMLVGFQWAGVLEQLDDGILNNSNKIFGHVKTENSTRATTEPVL
 NDRGTQFVLAFLVLRV

CPen_0557 642879 141194
cmcB 60kDa cysteine-rich GMP
EIMKGLIRKLVYFLVALPMAACGFAVYIIEAAVAESLITKIVASAEFTVDFNIVITAKKVF
ILVBNKQIVFQKIKGLAIDKDFQVIFCEGDFQVIFVQAQGVYGRLYSVKYNQKIVVEICQD
VPEYATVDFVDFVPELIIAGKKDVFVIVVITQDVFIAEFVDSPTPTTFIDPKLVWVLDIL
GADKDKITVWVKIKKGLKATATVCAACELRVTYTKKQCPQAICIKQVIGVACILGICPA
TVEIVAVYIAARMVTVNINVFVIAAGKQGVILFNLGDMRPIDKKVTVFVFGPDMG
QTNVATVYVGHKNGANVTTVLIEIVYVNIISADWYVCKVPEVYVGVNIGIDVILIL
DVVILQPLKPVATVLEAIVKIKKPKVHMKPQGVILFNLGDMRPIDKKVTVFVFGPDMG
TSEINQVTFSCAETITWKKGLAATIMVFLVNDPDCVIENTVTVIVCTNRDGEVIA
LILKGLPELQIAKATKPTKGLAATIMVFLVNDPDKVCKVSEVFGVTLIKLAPDAEGL

LKLLKNTPLTLHNTHTVINTLVNIGVGLDFARQF...
ITRLQYVPGDFVLSGSHGPOVTLGWPKFARKFFERL...
NRGLQGLKRIIFCSPPELCYITCSVD

Cpn_0579 669310 669993
ygbp/ncm-Gugar Nucleotide Phosphorylase
KEFASAPLLKATGTHVPMIKSLLSLGGQCTFRGSKI...
IPDIAEVIVVDDPQYOTFOCYVPSFAIPGERRQDSV...
LPHHILGISTAEF...
LAPFRAVITVDDTEGVELDSEKLVVNKHDP...
LAPFRAVITVDDTEGVELDSEKLVVNKHDP...
LAPFRAVITVDDTEGVELDSEKLVVNKHDP...

Cpn_0580 669936 670793
truA-Pseudouridylate Synthase I
ASSNQFLPRSSNDPSPMTKALLIAYOQTAYSQWQNDLS...
RTPLIASGRDAGVHAYGVAFHAPDFLANANLTKKALN...
FHARYLAIKEYRSLRLAKPLPQRHFCYTPRHPFSTEL...
HGRDYNSTVRTIYTLIDVIGKDSLSICRGNGFLYKM...
ILEQKNRRREGSAPAYGLSLHVVYSSPYNFCEQCSV...
ILEQKNRRREGSAPAYGLSLHVVYSSPYNFCEQCSV...

Cpn_0581 671533 670745
Phosphoglycolate Phosphatase
EGLRWRSVKSLFRCQWYVSMVLSDEFQCLRSQWY...
FLQACAEFLSLEHWFSTYSHHTLTGTEIFSKFIEQY...
PKAGPALMPGEAFIELVLSLNTKFGVNTSPTDHTLR...
PKYFGSDYDAYRTFAREGMKVIGFEDSVKGLRALSK...
GKEFFSYSPFDVLTCHSCQKLL

Cpn_0582 671305 672177
CT465 hypothetical protein
KPNALLKKIQRHRLVKMHDKNKLVLQANHLNKKRHN...
ISNVLKMLIRISTVSLTSCFSKNSRTCTVPERITSQ...
ISPNREVITAYSFYCGQGNISITPEGVYDCGLHHSIT...
DHPKVSITAEFCCKPHFLEASQSLQLHGTAAATFAL...
NDSPSLSNFIVTEATLNPRLRTQDGLSGHTEITVEL...
NDSPSLSNFIVTEATLNPRLRTQDGLSGHTEITVEL...

Cpn_0583 672349 672177
CT466 hypothetical protein
IVLSFFLGKTKTPRFLMNERLTLLLLKKKGLFLAIL...
IFLSCIDRVDLQKEFRHAFSSSELPODIEELEIRDI...
ERP

Cpn_0584 672659 673798
atgS/ntrB-2-Component Sensor
IRTNATHRRKRNLFVFNVPDSKHLHPAYELLEIKAR...
SECHFLICNSQAREILGIDENLEIRNSPTDVLPTD...
CKESKEKEVELFIRKNEISGVYFIQIRDRSDYKLENA...
NEESGIVGFASTLKKEISSPHORMLSISISGTRSLN...
DEESSILPLLSVSFPNCKPVREGAQFLRSIDPRMSV...
TSGBISVNTPGTIPSEIMOKLTPFTTFTKREGNGLA...
SEFTIPELLAALPKERAAS

Cpn_0585 675880 673865
*similarity to Cps IncA_2
ISIRKILRNPNFSIGDCSSNMPAKSPFQDPSFVREL...
IARVQCGWNTIIVKVSILIALITLGGSLVGLPAVPM...
LGLVDSQGLPEELFPVPEPQOQIEIDLNETREVLE...
QGMNTVRSSEGEKEISRLQDLISLQQQTVQDLRSR...
HDEASQACEGTMDCAERQLEKDLRLQKLSMQEWEM...
LEQDLRLTGLTAFDEQSLFYREYKEKYLQKLMQKIL...
QOYRSRLQEQKQVEKILEESNMHFDLPKAKQENMAY...
WLTDSASLSQKRIEVEENQELLKALAFKSNELTQL...
EQLRALDKHQAQIKDCEAAQRKCCDESILLSPVRED...
EQLRALDKHQAQIKDCEAAQRKCCDESILLSPVRED...

Cpn_0586 675993 677183
atcC/ntrC-2-Component Regulator
KEKNPSRGCENMAIKNLVDDPELRLDFSELLTSQGF...
LVDSMSMPDQSGDLIKIKQSSPHTPVLVVTAYGSI...
ALFAFISKAELKNLVHNFLLHSQTPDHSPLIAESKA...
GESGCKEVLFFIHNSPRANHPYKVNCAAIPTLLESE...
FELAHKOTLLDDEITVPVNLQAKLLRAIQEKEIHL...
IDDKSRQDLVYRLNVPPLHPLRLDRQDDILPLANY...
LNTYPMWPNIRELSNVLERVILENTSLTDEMLALA

Cpn_0587 677378 678124
*yvd_Bs conserved hypothetical protein
SYGELFILTLLKHVHTLQDKMRPHKRKHVSSKSLA...
ILEKSDHLPMTIRVLTSHKDKTEVHVAVSHGKEIL...
RTMANKHNSKRRDKRTDKGLAEEERIAIQEEDRLS...
PASAKKIKSKKMSIRMLSQDEATROLESANENFLI...
PSLKGPGCI

Cpn_0588 678033 678626
CT469 hypothetical protein
TSKSIKSNAFIKNMTATSLNLPSSQDSASESTQSQ...
LLSFLMHKLNYPKKLIIEKELKTLFLLMPKGTLPK...
GDGPKLLILECKALVNQALQKLLSYNYIGATC...
PGLPEYQQLNFIISLNL

Cpn_0589 678634 679395
CT470 hypothetical protein
GSMGICVTVVLRRLPLAKNHTLTFTPEGLTFFAK...
LHNNHGMPLKTIKIDILNAFEATQYVALLAEGKMK...
LHKEKEEQILQAIHAKQFSELLATAEFTATAEFT...
7HRIELPKVWPIY

Cpn_0590 680136 679516
CT471 hypothetical protein

LFLYGDHNLGFACPYLFF...
NMEIQARKKRVFNLITGE...
DFRGNWRSTITIGNQES...
YHNDLVGFSEVTLIFNV3SEGTITFS

Cpn_0591 680464 681020
yagE family
GLIMRCTAYCTASAYNLHVLPHLLKPRYPTILSREY...
LHMLRERIKMILITATA...
ANFILA...
HSDILDEPDDFWHDETOAIYRDVLSCLDLEARINVLIV

Cpn_0592 681132 681461
yidD family
LYSKMFSMSKFRLOQYVRIICLLIYLYQWILISPL...
LMGCLSLKRIKIGKCGHPOGIDMVPKALQEVLEPQ...
LMGCLSLKRIKIGKCGHPOGIDMVPKALQEVLEPQ...

Cpn_0593 682494 681391
CT474 hypothetical protein
VLGAKCMARKKRTKMLQVILSLVGLNMLFLFYSAI...
VYLSSEDFLNEISQASLDDLISLFDKDERVYGRPI...
TYLKGSSVRMLLBNIDLKDFPVILDVLRCHKYPTT...
CSTPEFLYRLTLVGADQVASSVASLARMVIRCGSE...
KSYLDCEESLALLLLVHDSVVLHEFCEDELEKVI...
LACMSTORVEAPRVQEDQDEYVVDGDSLWLAKR...
LKLPAKQS

Cpn_0594 682517 684958
pHet-phenylalanyl tRNA Synthetase Beta
NCHYTOYVVKSLVKTSLRLSMRIPITLLQTYFSEP...
LYSFASVITAKILHTIPHPNADKLRLVATLDGKE...
FDSEQAYTIKSKLGRVSESGMCGADELGLDELQI...
NTLSLEISLTPMLGCHASFLGLAREICHVTOANLV...
FSYVITGISAQSPSIKLOESLQALQKPINAVITD...
LRVKSITPESLTLNGETVLLPSGVVVRDDHSLG...
AYFLPLARASQKLLPIPSASAYRTRGIDPQNVV...
GEICRELKVALRPKTLQRLQKSFSEILSQKLSGL...
EIDVVEEICTESWNIEQNPVSCYTPYKLRKRETA...
LTKVKEEISLQGSKHTVLRSSLLPGLLSAATNLN...
ETQFLLAILLDEGSESWLPKPSLSFYSLGKQVER...
QOGLRIHKQSATLQGVHPELAKKAIKHPVFAEL...
VYRTLSNQDIEEYCRVALLNELLTDTKGTINS

Cpn_0595 684943 685926
CT476 hypothetical protein
RDYQFNKQLLFCVCFAMSCSAYASPRRQDPSVM...
LTKVKEEISLQGSKHTVLRSSLLPGLLSAATNLN...
LTKVKEEISLQGSKHTVLRSSLLPGLLSAATNLN...
LTKVKEEISLQGSKHTVLRSSLLPGLLSAATNLN...
LTKVKEEISLQGSKHTVLRSSLLPGLLSAATNLN...
LTKVKEEISLQGSKHTVLRSSLLPGLLSAATNLN...
LTKVKEEISLQGSKHTVLRSSLLPGLLSAATNLN...

Cpn_0596 685930 686457
ada-methyltransferase
FAVMAADTLIPKLMKNSLQACSEGLLIAKYPPLQ...
PGLPAHKAQEIIVLWCSRYANKEHPPFSHFAKOLI...
IAKATITDTRPTVGAACKNPFLFFFCRVGVSHGR...
IAKATITDTRPTVGAACKNPFLFFFCRVGVSHGR...

Cpn_0597 688215 686479
oppC-Oligopeptide Permease
MOKHPSFYQRFSLAYYKLLASLWKFFISVALIGI...
RYLDFPGYTKPVDFLFFNVMVTPFFILSFKTLR...
KVLQDPALEAKKMAEKVRENISKVNSEMMVLLPK...
KYRKKQEAQVKKYQVAFEEKRQSPMPTLRHLEMK...
WNRATDNYRPFMALTRIEHELRLADYNNWGPED...
LEDYAKLSAISFIQDKRLWIEKESDLRLINPFFS...
SRVTRKDLAALVFGIRIALVAVGIGITIALAIGI...
THRVFLTLMVLISITQKSLNLTLLVLCGFSWTF...
SHYVIMVHQLPNAIVPVISLPPFAMMAMISCEAG...
PAESAVLWPPAILTLMILLIALIGDGVRLDALDRLQDS

Cpn_0598 689712 688219
oppB-Oligopeptide Permease
EEGGSVLKYILKRLVLIPLTLFAIVSINFIINAA...
GKPDYQLQFREHYGLTLPFFMTBPKITHKKIQTAL...
DCAKVPMPALLFEADDSRDKYRIHAADLFIRGV...
AFVLRLNEEDLTKVEALKQWPDHGTQEVFCYSQ...
FOTLRNDAHKTIVSEVIRKLKCSLVLSLPMVGV...
FLILFSIPVFAVAVILNDVFNHTIPFTIPMPYSQ...
GFLPFCVSVYALAAQSRSLRSIFLEVLSDQFI...
TSLASSLGLTGGALVYVETLFDGPGNFYQALIN...
DICVLLDPRVQLEGRII

Cpn_0599 691823 689682
oppA-Oligopeptide Binding Lipoprotein
KRREGMHMKRCVLDKILKGVAGSLILLYWSSDL...
VKKQYTSQAIPAAGVMLAPKL/REAFALLPQDPS...
PHGILRTAHVCKDENLSPNCFD/VVGFYDLAIP...
EDGQZDEKFIYLRNVNWRPILKPLKPHVLDV...
ATMRAVALRXYEPVSVSVENLLKLVVWKAITV...
DREYVYFANGKEIIEDENDITPTNEDWQNTFM...
SRNIVYDPLAALIDKRFVYFETDLEPQKTKKID...
KQVAVYAVRETVADRAYTYLANKFELFEQVQV...
LPHACSPYNYKLEWNIETFEFAAPLEESWIDTA...
YVYKVTATITADYVATACKETIEEEL/ELAMAD...
FRLMWDJAMEKSAWVAFHIEEAFYIDUQVYED...
FLPDKESLLYKDVKNIVVTHPTLTHAQTIVNVM...
FLPDKESLLYKDVKNIVVTHPTLTHAQTIVNVM...

Cpn_0600 692174 691827 691827
No robust homolog present in GenBank/EMBL as of 11/7/98

GRPNFKSPYCLSDILYHNGKADKSTPAEVLALRSEHPIEKLAFRTIEKLLS
TYVKALPQVDSHTORHPSGAGVTRKACRPNLQNPPIEGRGILLRKAFRLSEK
NSYFLSADYSQIEARFLAKQDKSLKAFEGEDIIHATASQVHVHPVLEQSKQMGRA
TUNPFGIYVYQQAFLKVLKISICEAGELICAVFSRYPIAHFVTEIQQAAKDLRVTT
MLGRGRIIDSWNEPFGSRAAGSFAVTRIQGSAAELIKLMLDSQAIKQOQKMSRML
OHDHLELFPVEETIEEMORLVEKMESSMTLSAELPIVNNILKQKMAEC

CPB 0613 705662/ 704658

[illegible]

CPn 0614 707435 705783

adT-ADP/ATP Translocase
VIRHKHQGEFQMSSEVPPSRRLRAYLCPIYKSEFSKVP/LFLAFFVGFVNYLLKNMKD
TLTVVGGSDACAEVFPFVIMVGVIVGAVITVMYMGQVSKSPRYDTVYFCMAAFVGFLEFF
AVIYVPGVDSDLHNLSDAKLQELLQGLRGFRITVMVRYWSYISVYVMSLESSVFLSMLFL
GLANQITITTEAGRFALINTGLNLSICAGEISYWMGKQTFVAYSFACDSWHSVNLNLT
MLTCSGLIMWLVRIRHMLTITVTPSPSRRLAAEGEATANLKEKKPKAKARNFLHL
IQSRYLGLAIATVSNLWHLFEVDWKQDVQISYVSHFVNGPQVNTLIGVSWFLAA
VLLTQCKIRKQWVGFALTVPLMVLSGLLFFDKTIPAAKRDISFGVGLNLTPLAALAE
QGMQNLVRSQTKFVDTOTKENAIPLSPEEDKNHGKAAIDGVRSRIGKSGGSLTYOGLLV
IFSSVFAASNLVALVLLIMVWIAVVAYIGKEYISRAADAVATLKQKPESSIVYEAQ
ESVSEFMVAV

CPD 0615 / 708149 707634

pgsA-Glycerol-3-P Phosphatidyltransferase
LAKIMRQFNLLSLRSLWALALYFCQEKLIHRLLAIVGAMLSDDLGLYARRYKATSRGLS
ILDPIPTDQFVFCVITVLYMEGSLSI AHLFFICARDLFLIFVCYLSLVKWKGYDYGSL
FWGKIFTVQFVILILGVTAGGEIPWTGLVPLVLAFLGYFLERINDYKKOFLR

CPN 0616 708704 710137

dnab-Replicative DNA Helicase
 KLTTFVSSLLMDKSTGVPSPSPHSKSEMTVLGCMLTGTVHYLNLAANQLVEEDFYFYLEH
 LTITFVLQDAFKQDKPVDVHLAGEELKRHNQITVIGGSPSYLITLAEFACTAAYLEEYDVI
 IRSSTILRKMISTAEKIEKRALQEPQVNAELAEQDSFFKISQTSVSQYTLVLADKLRI
 LITDTKPYVLQIQERQELFLQNAQAGKCSLFFPTIGPDLQDLHGFSFSPNMLLAAR
 PAUKGKTALANIAENTLFCNQRPLPGIFSLSEMTVDQLIHMRICSRSEVDSSKISIGDLSG
 DFRIVSVINEMQCHTLIDDOQGLQSLDLARLRAMRSEYDQFLITLDYVLKLSGSGT
 RATESSQETISEISMRLTLARELNPICLQSLSRKEDRANHRRPMDSLREGSTEQD
 SDLVMLLLREYRQDPNDKPGTAEITAKNRNHSISGVPVLEFEKALFRFNYSAFECIS

CPn 0617 . 710481 712316

gida-FAD-dependent oxidoreductase
 LNMTHPIAYDVIVVGAGHACEAAYCSAKMGVSMVLMLTNSLDTIAKLSNCPAVGGIGKXH
 IYIEIDALGQIMAEVDVDSQVIFRILNQTGPAPRAQVDRVHLMKRLLENTPGI
 HVTMOATVSSLSDKEGVISGVTIKGEMSGKTVLLSGSTFMRLGLHIGDRNFSGGRGLD
 SSGSLSEDLKKRGGPIRSLRKTGTTPRLRLASGIEGFCMEEPDGLGVFVHREPPFPLP
 OFSCFIITHETKTKAISANLHRSALYGGCIEGVGPRVCYEDIKVFSDKERHNVFLP
 PEGHITQIEYANGLSTSMFPDQVQDMISGLVNLKNAITRPAAYIEYDIHGNVITHDTL
 SKITBGLFLGCGINQITGTYEAAAGCLIRAGLNVNENVPPIPSROESYGVMLDDPT
 TOILDEPYRMTGRAEHLRLSDYQDNACARLSHYGVELLGSIEERYLVKKONLLEEKVY
 RLQKTFPRYGQSVSLAKALSRPEVSDYMLREAFNDIRDLGAVNLASLMEIKYSYQI
 RMLITQSLEKASESLIPEDLVKQITALSLEAQEKLAKFTPTPGSASRSGIASADIQ
 VLKILKXKHAH

CPn 0618 712300 713010

lplA-Lipoate-Protein Ligase A

KNMPTTNCIFLDRGHSILHQLQIEALLRVANQNCFIINSQAKDSIVLGISRNLNQDVH
ISRAQADHPIPIRRYSGGGTVFIDSNTLVMVSWIMSSEASQAQOELLAWTYGYISPLLPN
TFSIRENDYVLGHKKIGNAGYIQRHRWHHTFTLWDLIDLKLSYYLPPIPQQQGYTPRNQR
SHEEFLTLTRPNPWFPSRDDFLERIKASGSLLTWEEFLDNELEEILAQPHRKATTYLN

CPn 0619 713462 713013

ndk-Nucleoside-2-P Kinase
RRYVYTMETQLTSIIKPPDSVSKAHIEGLISIFEQSGRLRIAMKMMHLSQTEAEGFYFVHRE
RPFFQELVDFMVSQGPVVLVLEGANAVSRNRELMGATNPAAASGTIRAKFGESIGVNAV
HGSDTLENAAVEITAYFFSKIEVNVNAKPLV

CPn 0620 714145 713519

ruvA-M Holliday Junction Helicase

CPn_0621 . 714707 714144

ruvC-Crossover Junction Endonuclease
LSRLGSSGFKDNKFKVFQESIVSELIGVDPGTIVAGYAIIVEQRYQLRPYSYGAILRSS
LMLPLMRYKTLFEQLSGVLDLDTQPNAMVLETFVNKNPQSTMTKLAMARGIVLLAAQQRDI
LIFEYAPNVAKKAVVGKGHASKRQVQVMYSKILNVPEVLHPSNEDISAPALNAICHTHVA
LSPALCGVR

CPn 0622 715761 714793

TYSYD hypothermical protein
 YVPLPLSLKELKELHSRASSLSLHYTHGRSRGMULLLCKWKDADMEWQOICNLSGV
 LPPMCKKLYLQKQETQDRINIEHERIHLYQYRQLTALDEEYPRREEAKHODLEKQQENT
 LKHLNARKLQQLRIQSDITDEIKKELLQKQETDISEPRLCYEYKIHOLEEQQLQRYYS
 LQHLNLEEDSKAAAYRINPLKAKLIDLQKQETIYKTYHIEIAKLRLQRBQQA
 LKQYRQYIEKTYVYDLEAKPKATALLQIVEDYQYQYLDIHKIKGMAMPDNTKLDHLK
 LKLEPESDQVADVEKELK

```

In [962]: %matplotlib inline
           %load_ext 2to3
           %load_ext 2to3
           %load_ext 2to3

```

TP94 hyperbolic protein
TPGTVVTVTRDVIETVITSPFVYKELVYRNTKHFQODFMVEALEVIVLNLICFFRNCQ
LSPFELVYAGDYSKMEVNRVTEIHLKAYGLDRAVKIAGCEALIKLKECTPLPLVIDEKPLA
EPEFVETDPPEYKKEKKKAEKLEKPKKKKKLSPRRIHKEIAEVTCAGDQELDPLVK

RELWEE:QFNEIVEQKFFLLPPPAKLISEVIGQTVADLNESLQALVRESSDL
INALLCAADUAHFPEETEPTSAFSESSAMFFPETS

CpN_0624 718018 717011

gpa-Glyceraldehyde-3-P Dehydrogenase

AMKVVINGFRIGRLVLRQILKRNSSVEVLAINDLVPGDALTYLFKFDSTHGRFPEDVRC

EADHLIVGKRKIQLSERVNOQLPWKDLGVDLVIECTGLFTKKEDAETHIOAGAKRVLS

ANWGDITPTFMYTNIHNTENPEKDFVIGNASTCTTNC LAPIAKVLNDHFIIEGLMTTVAH

ADNPAZTILFFPHWYFPTLDAIIPADNAGAYATLALPELRVLTMAFRVPIEDV

SVLLTFLRLDFTTTHADIKAMKASELDLRLVLTDEQVVDVDFIGNEYSIFDALVY

IALLNDRFPKLVAMVYDNETGYATRIVDLLEYEKNK

CpN_0625 718488 718060

r117-L17 Ribosomal Protein

VMQHARKKFRVGRSTSHNRCLANMLKSLIHIERIETTLPAKELRRHADMITLAKNS

LAARRIAIGRLMVRYNKLTSEARQAKGGDTSVYVDRVLVNLKDFELGNRFVERGGYT

RILKLNRIQDGAQKCIIEFLAS

CpN_0626 719670 718495

rpoA-RNA Polymerase Alpha

WLPAKKAQSVVLGFEKMGSDNAHLLYDKFELPEAVKMLPVEGLPIDKHARFIAEPLER

SGMHTLGNALRRALLIGLEAPAIISFAMTGVLEHMAIEGVIEDVTNIIINLKGALLIKY

PMQDSSLGRTTQVLKASISIDASDLAAANGQKEVTLQDLQEGDFEAVNPQVITVTOP

IQLEVVLIAGFRGYTPSERIVLEEDKGYEIVLDAAFSPVTLVNYFVEDTRVQGDTFDR

LVLIVETDGRVTPEKALAFSTQILTKHFSIFENMEKKIVFEAASIEKENKODLHLKLI

LGINEIELSVRSTNCLSNANIEITIGELVIMPERLQFNFPGKSLCEIKNKLKEMKLEL

GMDLTQFGVGLDVKEKNWYAEKIRAGYTKG

CpN_0627 720059 719640

rs11-S11 Ribosomal Protein

FLIRSLRVNQQAQKSVKRLQKNIIPSGVHVHVFATFNNTIVSITDPAGNVISWASAGK

VGYSGSRKSSAFAPATAAQDAAKTAMNSGLKEVEVCLKGTGAGRESAVRALISAGLVSV

IRDETTPVPHNGCRPRKRRV

CpN_0628 720461 720063

rs13-S13 Ribosomal Protein

DAYTILREAAQMPRIIGIDIPAKKLIKISLTIYIGISARSDEIKKKLDPPEARASELT

EEVGRLNLLQSEYTVEGDLRRVQSDIKRLIAHSYRQGRHLSLPVRGQRTKTSRT

RCKRRTVAGKKK

CpN_0629 721881 720487

secY-Translocase

KIIEFRPYMTTLRQFLITELRQKLFYTFALLTACRVGVIPVPGINGELAVAYFKOLLG

SGMFLQADIFSGGAFQMTVIALGVVPIYASIIIVQLFLVMPALQREMRESSDQGR

RIEGLRLFTVALAVIQSLFAKALRMLNLTIPGIVLPTLLSSKLGFPWIFITTVVM

TIGTLLMIGEQISDKGIGNISLIIALGILSSFPVLSGIVNKLNGSQDSSDLGLIS

IEELAVFVFLITITILIEGVRIKIPQYARRVIGRREVPGGSGYLPLKVNAGVPIVIF

AGSLMFPATIQGFIAESSEWMMKRIALLAGSLVYSICYVLLIIFFTYFWATQFHPHQ

IASSEKKNNAFIPGIRGKPTQHYLEYTMNRVTLGALFLAAIALLPSLGLCLLRVDSNV

SYELGGTAMLVVGVGLDTMKQVDAFLLMRRYDSVLKTRDRTKGRH

CpN_0630 722316 721885

r115-L15 Ribosomal Protein

MKEGESLFDISERKRRLKLLGRGSSGHGKTSGRKHGKDGSRSGYKRRFGYEGGVPLVR

RVPTRGFSKHKRDKCEETITGRALAEFGGEAITLDALKAKKAIARQAVRVKVLKGLD

EXTFVWDQTAVALVSGVQNLGIT

CpN_0631 722812 722312

rs5-S5 Ribosomal Protein

EBMSLSKSHKEDQLEEKVLVNRCKSVVKGGRKFSFALILVGDGKGRIGYFKAANEL

TDAIRKGEAAKNLMKIEALEDGSIPIHEVLVHHDGAQLLLKPAKPGTGIVAGSRITLIL

ENAGIKDIVAKSFGSNPNMNVKAAFKALITGLSPRKDLLRRGAAIND

CpN_0632 723354 722827

r118-L18 Ribosomal Protein

KGLTSSWLNNLQVAPNVLNLIKREFVMMKMSVVKLVLRKQAKNRSRPMESSLCK

KSEIKRRRLRVKLVKGSPTKPLRSVVKTNKHIYVQLIDDSIGKTLASVSTLSKLNKSG

GLTKKQEVAKVLTQIAELGKNLQLDRLVFDGRPFKYHIVSMVADGAREGGLOF

CpN_0633 723760 723209

r16-L6 Ribosomal Protein

SMRKAKEPILLPGOVEISIDDKIIVKGPGLSTOKSVKEVEITLKNSIFVHAAPHVV

DRPSCMOGLYALISNMVQGVHGFKEKLEHIGVGRASVQGAFLDLSIGVSHPTKIP

TLQVSVKNTLISVKGDLQKLVGEFAASIRAKRPPEPYKGGKIRVENEYVRKAGKAAK

TGKK

CpN_0634 724215 723787

rs8-S8 Ribosomal Protein

ESSIKRRIYKMTSDSADLTLTRINALMAEHLYVDVHSMREAIVKILKHKGFVAHY

LKKEENRKRMRVFLQYSDDRKPIHQKRVSKPSRRVYVSAKIPYVFGNMGISVLS

QGVMEGSLARSKNIGELLCLVW

CpN_0635 724763 724206

r15-L5 Ribosomal Protein

GERKANMSRLKFFTEIRKSLFEKFGYANKMQIPVEKKIVLSMGLAEAAKDNLFQALH

BEELTMISQKPLVTKARNSIAGFKLEGGGIGAKVLRGIRMYDFMDFCNIVSPRI

DFPQSNKQDGRCCYVGLDDQIFPEINLDRVKRTGLNITVMTVTAQTDDECTTLELMCL

RFFKAQ

CpN_0636 725100 724754

r124-L24 Ribosomal Protein

FKEKEVMKKNIIRVGDYVFLAGNDKQKEGVLSLSTEDKVVVEGVNVRINKIKRSQONPK

QPRIGIEAPTHIISNVLITAGFAKLGVKYTEQRELMQRRPFTSGLYLRVKGKK

CpN_0637 725473 725099

r114-L14 Ribosomal Protein

IEIMIQKQKQKVAQNTAKKVKFKYLOGGRRRYATVGVIVGVVDRVDEHNSIKKGV

IVAVIVRTHHITKDKSTLKPDTHACVITDDKGNPKFTRIPGPVAREIDRGRFKISGL

APAVI

CpN_0638 725774 725490

rs17-S17 Ribosomal Protein
NKKKVKSMAGEPPGSRVKVLSAKMEKTVVVRVERIFSHPPQYLVKVRSSKKYYAHT
ELKVSQEGDKVKIQETRLPSKLKRRVIEHVGVVS

CpN_0639 725979 725743

r129-L29 Ribosomal Protein

ASGGKINMAAKKDLTLQRLGKSDDDLDVAVHKNKALFALRAENLLQNKVVKHMFSTHK

KNIRALTVKQEPYGVKHG

CpN_0640 726049 726409

r116-L16 Ribosomal Protein

IIMLMKRTKFRKQKQFAGLSKQATFVDFGEYAMQTLERGWTSRQIEACRVAINRYL

KRRGKVVIRIFPKSVTKKPAETKAGKGGAPDHVAVVRPGRILFEVANVSKEDAQDAL

RRAAKLGIKTRFKRVVERV

CpN_0641 727092 726409

rs3-S3 Ribosomal Protein

KGRIMQKQKCPIGFRTGVTKWRSLWYGNKQEFKFLIEDVRIOFLRKKPSCQGAAGF

VVRNMGKLEVTIQTARCPVIGKGAEDVLLKEELRALTGKVEWLEIAEIKRPELNKL

VADNIARQIERRVSFRRAKKAMQSVMDAGAVGVKIQVSGRLAGAEIARSEWYKNGRVPL

HTLRADIDYATACAEITIGIGIKVINLGENSSSTTPNNPAAPSA

CpN_0642 727440 727096

r122-L22 Ribosomal Protein

RRHSMFKATARYTRQPRKARLAAGLMNRLSVQAEQGLGFSQLKAGRCLKXVLSAVAN

AELHENIKRENLSITEVRVDAGPVYKRSKRSRGRSPILKRTSHLTIVIGEKER

CpN_0643 727725 727450

rs19-S19 Ribosomal Protein

EIRIMGSLKPGPFVDHLLKXVRANIEEKKTPIKTWSRRSMITPEMIGHTFEVHNGKK

FLTVFVSETHVGHKLGESFPTTRIFKSHPVKKG

CpN_0644 728594 727722

r12-L2 Ribosomal Protein

FIREINSMFKKPKVPTGTRQLVLPADFELTIRGELRGTKSKRSLRPNNKLSFFKSSGG

RDNLGISCRRHGGGAKQVYRVDFKRNKDGITAKVVTVEYDPMRSAYIALLSYEDGKRX

YILAPKQIQRGDVVSGGSPFKPGCCMTLKSIFLGLSVHNIEMRPSGGKLVRSAGLAA

QVIAKSPGYVTLKMPSGEFLMLNEGCRATIGEVSNADHNLVRDQKAGRRMMGVRPTVRG

TAMNPDVPHGGGEGRHNGYIPRTPWGKVTGLKTRDKNKNKWKVDRRK

CpN_0645 728933 728598

r123-L23 Ribosomal Protein

DMKDPYDVIRHYVTEKAMLEHLSAGTGEGKKKSGFCKDPKFVIVSHDATKPLIAQAL

EATYVDKNVKSVMNTINVKPPQARMFRGRGRKTSFGKKAIVTFYQGHVSG

CpN_0646 729636 728950

r14-L4 Ribosomal Protein

YREDMLVLSKDFSGNKGVEVADSLFADEGDLQLIKDYIVAIRANKRQWSACTNR

SEVSHSTKPKPKQGTGNARQGLASPOFRGGGIVFGPKPKFNQHVNRINKERKAAIRLL

LAQKIQTNKLTVDVDFVVDALTAPKTQSALRFLDCMNVCSRILFIDHLDHVEKNENLR

LSLRNLTAVKGFVYGININGYDLASAHNIVISKKALQELVERLVSETKD

CpN_0647 730490 729657

r13-L3 Ribosomal Protein

YLEYFSYCKNLPLITCPFIPLRENFLFLENISKILSRFVSLFLOEESKLLMDKFM

RSHISVMGKMGHITHFDKQSLVACSVIRVEPMVATQIKTESDGYFSLOIGAEENAP

AHTITKRVSKPKLGLHRLKAGORVFRFLKEVRGSEALNGVSLGDAFGLVEFEDVSSVDVR

GISKQKPGQVMKFGFRGQSGHSGFHRHAGSIGMRSTPGRCFGSKRPSHMAENVT

VKNLEVIKVDLEKVLVKGAIIPARGSVITVHSSRT

CpN_0648 731636 730605

Ct529 hypothetical protein

FFFKKPKCEVMQATNARSAGSAASKMLLPVAKPEAAVSSFAQKGIYCIQOFTNPGNKL

AKFVGATKSLDKCKFLSKAVSDCVVGSLEAGCTGDALTSARNAQGMKLTREVVLANV

LVAGVAVINVTORCYQYTRQAFELGSKTKERKTPEYSGKMLLTRGDVLLAASREACTAV

GATYTSATFGLRPLMLINKLTAKPFLDKATVGNFGTAVAGIMTNMAMVAGVAGGIAL

EOKLFKRAKESLYNERCALENQSQSLSGDVILSAERALRIKRVNLTLEKALEL

VVDGKLIPLPITVACSAISGALTAASAGIGLYSIWQTKSGK

CpN_0649 732672 731710

fnt-Methionyl tRNA Formyltransferase

LNKVVYFGTPTFAATVQLDLHHKIQITAVTVTRVDPKQKSAQLIPSPVKTIATLHGLP

LLOPSKASDPQIEELRAFNAVDFIVVAYGAILRQIVLDIPRYGYNLHAGLLPAYRGAA

PIORCIMEGATESGNTVIRMDAGMDTGDMANITRVPIDPMTSGELADALASQAEVLK

TLQIESGQLVLSQDAALATIAPKLSKEEGQVPMWPKAKAYAHIRGVTPAGAWTLFS

FSEKAPKRLMIRKASLLAEAGRYGAPGTVVTDQELATACSEGAICLHEVQVEGKGSTN

SKGFLNGYPAKKLIVFTLNN

CpN_0650 733513 732665

lpxA-Acyl-Carrier UDP-GlcNAc O-Acyltransferase

SRNMASIHPTAIEPAGIKGDVIEPYVVIKATVTLCDNVVKSAYIDGNTTIGKT

TIWPSAMIGNKQDLKYQGEKTYVTIGENCEIREFAIITSSTFEOTTVSIGNNCLIMPWA

HVAHNCTIGNNVLSNHAQLAGHVQGVYALQGMGVGVHGFVIGAHAMVAGALSIRRDV

PPYITGSGNPYQLAGINKVGLQRRQVFPATRLALIKAFKIKYRADGCGFSELETLEEYG

DIPVKNITIEFCQSPSKRGIERSIDQALEESADKEGVILIES

CpN_0651 733975 733517

fab2-Myristoyl-CoA Carrier Dehydratase

MNQPOVILKRELLDPLPHRYFLLVQKLVSYDIEARSITAKQNTINEPFFMGHFNAPI

MPGVLLLELAQAQVVLGLVLEHDKRILALFLGIQKAKFRQAVRPGDVLTLQADPSLI

SGKFKAWAQARVDSQIVTEAELSFALVDKESI

CpN_0652 734880 733990

lpxB-Myristoyl-GlcNAc Deacetylase

KRNLITVLDGLSPYMLERTORTLKREVRVGVGIIHUKSTLHLQPAQNTYIVFQRQ

ASVETNVPALLPHVYTNIRITLSSGCAVATVEHLMALRSNNIDNLIITQCFEETPI

QDGEHNVVELIDAGICEQEDKYSIARLTPPYVQHQDIFLAAPPDDELKITYLHYPO

STPTGYTKGLVINEECPROELACPTALVLELCLFMEKLGICQGLDNNAVVPKDGII

SRQALRAFEPVRHKLDELGLDGLVGRFVAVHVLAVGQGHSSNIAFGKKILEALEL

CpN_0653 736499 734868

curE-Apolipoprotein N-Acetyltransferase
GEPVLRIFCFVIGWCLTAFAPDLSGFVSLGAACGY...SLEPLKKPSLPLRTLFVS
CFPMITTEGIFHFWMLSDQYIGRLYLVLWLTITLSLVSFGSCLLVAIVRQKRTAF
WSLPGVWVAIEMLRFGYFGMSDFVLGWPMTASAYGRQFGGLWAGQSVFAIVNMSF
YCLLLKKPHAKMLVLTLLPYTFGAIHYEVLKHAFOQDKRALRVAVQPAHPPIRPLK
SPIVWVQLQLVSPITQOPIIDLLIFPEVWVPPFGKRRQVYFESCAHLLSSFAPLPEGKAF
LQNSDCATLQHQHFCQVITGLERWVKKENVLWYNASAEVISHKGISVGVKRLVPGGE
YIPVKKFCTIIPOLFPYVIA/CKRLPGRRGTA/VRGLPRIGITITTEFTFGYRLOSFK
...
KEIR

CpN_0654 737051 736503
vdID/yCIA-acyl-CoA Thioesterase
KKIDFLSVDRYRNQEQYPIKILSVESTMLKKPVFSCIDGHIYKIFPNLNNANVTFG
GLLMSLLDRALVVAERHTESVCTAFVADALRFYAPAYMGENLCKAAVNRWRTSLEVG
VKVMAENIYQERRHITSAYFTFVAVNEDNQIPVHQIVPETPEKKRYNEADRRRQARL
ELK

CpN_0655 737856 737101
dnaQ-DNA Pol III Epsilon Chain
KEIMSLKDTVTFLDCDMTGLDVKKRIIEIAAVRFTFDSVISSIEFLINPERVVSAS
QRVHHISNAMLRDPKIAEYVPOIKAKFKEGDYIVGHVSFGFDLQVLAQEMERIGETFLSK
YTIIDTLRAKEYGDSPPNLSLESLAVHFNVPYDGNHRAMKDVENINIFKHLCKRFTLE
QKQVLAKEPKIMKYNPLCKKHKRCFSEIPLAYLQWASKMDFDSDLLFSIRHEIKHRQGT
GFSQVNNPFML

CpN_0656 737842 738048
No robust homolog present in Genebank/EMBL as of 11/7/98
THNLLLPLSLFDILLTVEGLCLTLFASVORMPEQKRVPGNLYYYIAHSSCLLSV
CKDTMENKD

CpN_0657 738476 738051
yjeE (ATPase or Kinase)
PMGRYRVSHSSQETLLGLYDQVLVPGAVLLFGDYGAGKTEFVRGIVSYGLDITAE
EVASPSFSLHYGNEPKRLCHYDLYRIDQKQVEYIFQDAEEDVLCIEWADRLPKPRFC
DTINITYITMQTMEREIIEKR

CpN_0658 739180 738455
CT538 hypothetical protein
KRWGMDISGAVKQLLQFGLGKQKPELLATYLFYLEQALSRLPVVFRDKIKFTPEDAV
RIEEDQDKIWRETEIISSEKPOVNMETKRIYICPFTGKVFADNVANPDADYIWLSSC
POMMEKQGGVRIKRLVSEDDPDIKEYAVPPKEPIIKTVFASAITGLFHSPLPLEDFI
SSELRPHMLEEVQNTKQFLESSFLSLQDALVEDKIAAFIESLADDTAFHYIISQWVD
EET

CpN_0659 739482 739838
trxA-Thioredoxin
LQENRDSNSIFREGKLMVKIISSENFDSFIASGLVLVDFFAEWGCPRLMTPILENLAA
ELPHVTIGKINIDENSKPATYEVSSITPLILFKDGNVAVRVLGDKKEFLTNLKNHA

CpN_0660 740327 739860
spcB-rRNA Methylase
MRVYLHCPDIPONTGNIGRTCVAGELILVRPLGFSADKFKVRAQMDYWDKLQTVVD
SFEALHWDVPEQIFCLSTKGSAYSESLPSGGTYVFGSESKGLPKEILKYYKNCLRI
PMQODIRSLNLTASVGLVLYEVVRQKTVALQKNPTV

CpN_0661 741139 740327
mip-FKBP-type peptidyl-prolyl cis-trans isomerase
HSRGLKIDRRRRKNNRWNLVATVALASVASCDSKDKDKQGSLSVEYKDNKDTNDI
ELSDNQKLSRTFGHLLARQLKSEDMFIEAIVAGLQALVCKSAPLTETSEYEKMAEV
QKSWFEKSKENLSAEKFLKENSINAGVVEQPSKLQYKIIKEGAGAKISGKPSALLHY
KGSINGQVSSSEGNNEPILLPGQTIPGVALQMGKMEGETRVLYIHPDLAYGTAGQ
PMSLLIFEINLIQASADEVAAPQEQNGE

CpN_0662 742938 741172
aspS-Aspartyl tRNA Synthetase
SKGQYMKRTHRCNELTNSHIGENVOLAGVHRYNRHGGVVFIDLRDRFGITQIVCREDE
QPELHQRLDVARSEWLVSRGKVCPLAGMENPLATGHEVEVASFEVLSKSNLPFSI
ADHINNVNEELRLERYLDMMRGDIIEKLLCRHQVMLACRNMDAQGFTIIVTLGKST
PEGARDVLVPSRIPGKFLYAPQSPQLFKQLLVGGLDRYFQIATCFRDEDLRAQROPEF
AQIDIEMSFGDTQDLPIIEQLVATLFAQGTIEIPLAKMTYQEAQKDSYQDKPLDRFD
LKLKDCRDYAKRSSSIFLDQLAGHGTGFCVPGGATMSRKQLDGYTEFVKYVAGMLV
WIKNGEGKVASNIKAFMDEEVFHELFAFYDAKDQDILLIAAPESVANQSLDLRLRIAK
ERELYSDNQNFVMTDFPLSLEDQYIAEHHPPTAPLEEDIPLLETDPLVRSSSYDL
VLNGYEIASGSQRHINPDQSQIFTILKISPEISQIEKFGFIALSGFTPSHGLIALGLD
RLVMVLTAAESIREVIAPPTKQASDLMMNAPSEIMSQLKELSIKAVF

CpN_0663 744220 742901
hisS-Histidyl tRNA Synthetase
KSNHFERRRHVTITLPGKGVDFIPYLADAKQLWRHTSLWHSVEKAIHYVCMLYGFCERT
PIFEKSEVFLHVGEESDVVKEVYSPIDKRGSRMTLRPEGTAAVRSFLEHAGHSRDNK
FYIILPMFRYERQAGRYRQHOGFVGAIVGRHPLRDAEVLALLWDFYSRVGLQHMOTQL
NFLGGSSETRFRYDKVLRAYLKESMGELSAQQRFSTNVLRILDSKEPEDEQIEIQAAPP
LDVSDDEOLKYFNEILDALRVLEIYAINPRLVRGLDYVSDLVFBATTTFQEVSYALGGC
GRYDGLISAGFASLPCAGCGGLERATQTLAQKRIEPOFPHKRLIPMEPDADQFCE
WQHLRLRIIPTVDMSHKKVKGALKAASTEQVSF/CLIGEREISQQLVTKMNSLRKEF
POTKEEVQRLLVEIQNTFL

CpN_0664 744775 744557
No robust homolog present in Genebank/EMBL as of 11/7/98
LWFAHMKLIALIGILPLVPIKGNTEKHDAHATVLAARAKYNLFVQDVFPVHEVIEP
LQDLVLIYEDW

CpN_0665 744998 746365
hpkA Hexosephosphate Transport
KMMWIKFPFPPKHLKELEQEVVKKYKWRIRIFCMFICYIYYFTKRSFTFAMPTL
IABHFKPAKALITITLTYFHYIGKFGVMSDQIPRYFMAICLMTGLTNIFFOMSS
SIVLPALWJLWQWQWNPICARLLTWAKDEFTWWJWVWJSHINICGALIPILTCF
IUYXWQWAMVYITLCLMGLVILNLRDTPQYILPPIEYKRDPHAHIEGKSASE
GTDEIRGELTFRILFTVYVITNGLWIFLAAAFETIYPMVAVNDWAGLFLIETKIHYAAVK

ANFCVSLFEIGGLFQMLV...KISKVNGPMMVLFSLQLLFAILGMWFSRHHQWV
DGTLLFVIGFLLYGPOMHIGL...AELSHKKAAGTASQGTQWFAFGATFACYPLGKVTDV
WGWGFFIALLACASIALLLFLFTWNAETKNTSKA

CpN_0666 746379 750107
dnaE-DNA Pol III Alpha
GFFLTWIPLHCHSQSVSLDAMSSIKDFVAKQCEFGIPALALTOHGNLYGAVDFYKECTOK
GTQPTIGCECYIAPSRFDYKKEKSRRAHHLILCKNEQGYPNLCILTSLAFTBGFYFF
PIIKNGILPLVYSEK...
...
VFTQDNERSVMSPPDAFVCEDEELRQELVADLPVRVTV

CpN_0667 751097 750177
No robust homolog present in Genebank/EMBL as of 11/7/98
NISLLCKIQRYFMKLLILYFAAFVSLFCVGLWDRVPCAQKIMRLAADHSSEVTSKSC
RFRVKISGFEEQLQVFERHVSPEQALALFPEYRDGKSFVELAIPHTLMHVRFSKEEPVKK
HISIQEGSILWSLVNGEMVLTGTWTCNSGRCFELLHLHAGQKDRVITQLALTGGTTSRE
SLAQUALKINRAERVIKCEQKKLIFASGQIQHTFQFQPIRGCTTTLANNPNWLPK
RHAAYFPAQYSEDRVRLVQMIFGDNFLIVRSMSVYVYKISLVSADNSRVEYINAVT
KGSFQBL

CpN_0668 751176 752162
CT547 hypothetical protein
WRVWVSPRLIMKFLLYVPLLLVLVSTGCCDAKPVSEFPGSKLSTORFEPQHSAAEYFSQ
GOEFLKNGPRFKALLCFGIITHHFRPDLIRLQAQYILGVYFTQDHPDLADKAFASYLQ
PFEYSSEELFQMYAIAQRAQAGKRRICLEGFPKLMNADEALRIYDELTAFPSKDL
GAQALYSKALLIVNDLTATKTLKLTQLFPLHLSSEAFVRLSEIYLQAAKKEPHNL
DYLHFAKNEAMKQHPNHLNPNVSVANVAMREHYARGLYATGRFYEKKKAEAAINY
YRTAITNVPDTLLVAKCQKRLDRISQHTS

CpN_0669 752140 752775
CT548 hypothetical protein
IYELSLPKEINMRLVSLGTIYLFSLALSSCCGYILNSPYHLSLGLKSLQERIFIA
PIKEDPHQCLCALTYELSKRSFAISGRSCAGTYLKVELLGIDNIGTFTYAPNKLGD
THRHPIVSNBGLSLSAKVQILNNDTQVLLIDQCVARESDVDFEPDGLTANAEHALQ
FEMHSEAIKSARRILSIRLAETIAQVYVYDLF

CpN_0670 752738 753196
rsbW-sigma regulatory factor-histidine kinase
PRRLNRYTMTTFEGEYFPAVLSLHSLMDLIRAKAGQSKCPOEKLKLELACEELLVN
IISYAYQGENSPOTIAISCSHRGLVVKDGHSPFNPLAVSINIQEDLPLQRLKGLG
GIFLAKSSVDEFLYAREDHNCIVHLKMLNGQHS

CpN_0671 753660 753205
CT550 hypothetical protein
RTIINORKYMSLDFFEYFQHSILNTQGSFPEGYNLAIILSYPHCTDANTDFLCSQSD
NDFIITAESKDKLTLNADFAIWLVLVQCAAVTRGYIAVSQGEQNEPEMAFEASGQYN
QSSLLLEALQYLDKIDKTENALRSFRFNNDH

CpN_0672 753723 755048
dacF(ppb5)-D-Ala-D-Ala Carboxypeptidase
TIKSPHMKRPFFTYLCIIFYGSCASLSHAGLSFPEVRGATAAVHADSGKVFYDKDIDA
YIYPASTMKIATLALFKHYPTVLDLTKVQDAIASITPOAKQSGVRSPPHMLTEDGS
TIQLHLREELLGWDLHALLVCSANDAAVLAACSSVEKFMKRLNFFLEIGCTHTH
FNNPHGLHNPHTYTTDRDLSIMRCLAKEPPRGVISTTSYKIGATNLHGERILSPNTKL
LLPGSTYHYPPALGKGTGTTKTAGKNLIAAEKNNRLVLTATGYSVGPVSDLVQVIALC
ETVFNPEPLRLKELVPSDCQLLEIANGLKSLCPLEGLYDYFYASEDEPLSVSFIAD
APPIEGQDGLLGHVWFYDDEGKISSQFPYAPCRFERTIKPWKLYMKRVFTSYRTYMSITM
LMYFRIRKHKRYKNLKHYSKI

CpN_0673 755242 755463
CT552 hypothetical protein
GKSTEGKAYHCFLKQVSIALNREEVDNPHMLFILMQFQFSGEQDRFSGLEATIRDR
VSFLVLQEKIATLK

CpN_0674 756689 755577
fmu-RNA Methyltransferase
RGILYVTPFPROHHAYQLLQKLTSAISEADRVSYFFKQNRSLQSKDROWQNIIFNIL
RHRLLTLLDLSGBQVTEALVAKVNEGVLNDSYSAIPWPVYSISDOLAHFLVQDY
GEOAEETAKIWLTEAPITIRVMTDKIGVKELQEKLEYPSSPGLPEALHFKSRHPLQST
EAFRRGFEIQDENSORTISQGLDLDKDIVLDFCAGAGCKSLIFAKAKHVVINDSRKA
LQTAHRLRLAGARFSLADQLRGSFVIVVADAPCGGTGVRRHPEHKWQFSKLLLLY
VRVQSLKQASAYVGRGLVYITCLLKEENEAHYVMHSLWKEVHRKTLQVQVKGK
DAFFTSHFKI

CpN_0675 757931 756788
CT556 hypothetical protein
VPLMILDFQPSIGYLRVLELAIRATTHIAYIRKRLILDWIVNIMLITNYDTSVSTI
QVHELFPWASISYSISRLALIELPHEKEITQWVYLFPEFTKYHKKAIQVKLM
FKGLILFESKRPVDKVOAANK/FGKPGNNSWEDTHE/TVSEVYTPIAEVRORLAA
DACQMIIEALTTLLEGTAYLLELILNOK/HEKQAIKTELKEK/VIILRELQLFSL
SAEDPOTIMSIDLSEVLALZLQDIP/PIIKTFTY/MLQRTALAEDEKIALGLF
AEVLRKVIYEKKLHVCKNWTTFEEVCH/IT/IN/NTALMDKMTIMIIMWILDYDRDGL
IALKAAEYINPHSPWRQFLPLQWPPF

32

TADSLAKDIAMHIVAAQPOFLSKESVPAEIAKEKEV...GKPOEVIEKIVTGKLT
FFQACLLLEOPFTKNADLTQSLIDDFGKTGGSSVA...KIGA

Cpn_0649 783443 784201
pyrH-UMP Kinase
EPNKNMAKQTRRVLFKIGCEALGKSSNRIDEMRLSRLVSELRAVRNNDIEIALVIGGGN
ILRLGLAEQELQINRVGADOMGLMATTINGMAVADALKAEDIPCLLTSTGCPQLADLYT
POKSTIEALDOKILITCTTAGGSPYLTITDGAALPACELNDVLIKATMHVDGVDKDRPL
PQVATYDFVAVVHDLTLPJLPAAGALILMAGHIFIRVFPJLJLDFALDPTITTL
VAVLVNHHVH

Cpn_0699 784179 784721
rrf-Ribosome Releasing Factor
TMSVLQDTEKKMAAALDFHKEVKSFRGKAHPALVETVVVDVYGTTRMLSDIASISVAD
LRQLVISPYDGNASAIAGKIIAANLNQPEVEGSIIRIKVPEPTADYRQEMIKQLRRK
EAKINVRNIRREANDKLKSDALTEDVVGNEKKIQELTDKPKQLDELTKQKEAETAS
I

Cpn_0700 785094 785609
CT676 hypothetical protein
LMVHSPHQYCHQCPATICYTEIDKDKVIRSYVATCPSPSHYNNNEHLSLKGVGVL
LEGCKTVMHSGKQDDEQLGCHQCYTNFKNQITSKLKSERVSSFTMEKGGSLHIGR
APGEASNTNPLKLIALNEALQDTEREDYEQAIVIRDQINHLTKNPPDPS

Cpn_0701 785584 786672
karG-Arginine Kinase
KPKIQMTLPNDLLETLVKRRKSPQANKVMPVTFSLARNLSVSKFLPCLSKQKLEILQF
ITSHFNHIEGFGFIVLPKDTPLMKQEFLEHFLPYDLVGNPEGEALVSRSGDFLAA
INPQHLVLHGIDFGQNVKTLDOVLQDLSYHSLKSFASFSEFGFLTTPNPKNGCTGLKS
OCFLHIPALLYSKFTNLIDEEVEIITSSLLGLGVTFGPNIVVLSNRCSLGLTEELLSS
LRITASKLSVAEAAKRLSEENSGDLKILRLSGLLTHSCQLEKETLDALSWQLGI
DLGLIKVTENHPLWNPFWQIRRAHALQKQAEASRDQKQITSLHRSVSKELTKGLSP
ESF

Cpn_0702 789700 786929
yscC/gspD-Yop C/Gen Secretion Protein D
LKNPVKTVILNIGRKILQGIKKKKIKIGLSGLFFLDLVLLGVSSQRPETSTANVKNL
RDEKLAACPKNSAASLAKSHTKTTPGSIKPVSKFDFATQDKTQKTSQSAFPAKPT
TLKLEERKKPRPERRTADVKRSPRLPTQVEEVPVPAASKEQLDSIQWKEQNVARR
AVNAINLSIKKQLEEQTSYTEKDVQPKTOATPHASKKNVSPSTSMFGIEKAATTAVP
QDKSEEEKVKERLTRELCTEDLKNGYTNFEDISILELLOFVSKISGTNVDNDLQ
FNATVSHSDPTSDDLSTILLQVLMKMDLVVEQNNVLYRNPHLSKLSVVDTSLSKE
TCFAVTVTRVRLSVSPSAVNIIQLLSHDAIVSASEATRHVVISDIAGNVKDDLL
AALDPCGTSDMTVEYKYNAPALVSQYDQVGLTAEADAFQMFIPQGTNKFVVSPP
LAKKAEQLKSLDVPMAHLLDDPASTALAGTGTTSPKSLRFPMYKLYQNGEVIANA
LQHGYNLYVTAMDEDFINTLNSIQWLEVNNSIVIIQNGQNVDRVIGLLNGLDLPKQV
YIEVLILDTSEKSDFGVQWALGDEQSVAYASGLLNNTGIATPKATVPPTGPNNGS
IPEPTPGQLTGFSDMLNNSAGLGIIGNVLSHGKSLFTGLGLLSALQDQDVTIVLNR
IRHAGDQQAASFPVQGTVPYQNTINTIQTGTVTQNDIDYEDIGNLVSTVAPNVVTL
QIETITSELHSGSLTPVTDKTYAARLQIPDGFVMSGHIRDKTKTVSGVGLNLSI
PBERGLFSRTIDQRKRNIMFIPKPVSISSFEGTRVTNKEGYRNWEADEGSMQVAPRH
APCCGPPSLQAESDFKIIIEAQ

Cpn_0703 791205 789685
pksA-S/T Protein Kinase
RKIGFMDCRGGIPLPEPOVIGYHVKILSKLRSRVVHGLHPETRHSTVIKVFSPSEF
TSRSYVNFLEKQASQLTHPNIVKPHRYGKQDCLYIAMEYIEGSLREYILAQFISLP
QAIDIIFDIAQALEHLHSRNILKDKIPENILITPOGKIKLIDFGLADWDEIQRAPSV
IGTFYPMSPQROGESHPASDIYALGLAYELILHLSLGRVFLSLVPERISKLAKAL
QPSPNRYVSTREFIQDIHHYRMSQDQEDLRIKDHVALYEQLTQRTORFLAPETLRFPD
FISGVLVYHQGPPLYPHAYDTLLEGGDVFNWLGYSPIANATIALSVKSVLCCQDLRPL
DRVEIKINECLIRMKIPIDEMGISILCLISKENKELSWIACCGTIVKNGKVVQDFES
FSPGLGKITSLQIRETKVAWEIGDEAVVCTLEESVASLKTLSLAELQDRAQAKIFCPI
ESHGQISQRQSGNSPSTLISLKRIR

Cpn_0704 792330 791209
fliN-Flagellar Motor Switch Domain/YscQ family
RYEMAAVADSSASWLKSRNNFLSSLGKTEQVAAPFPKELQCHKIPEKFRLEDVQVSIK
FRGSIATAVEATKEFGVHLLIQPMVQVPEVENLLFTSEEDQELVAVFDASLASYY
EKDKLLGFHYFVAEACKLFEELQWVPSLSAKVGGDAITATSLQSFQVVDISRLDGK
NVRCLLLPEDTFQSCQKFPGLHDSQDLHNDIQTOQISLSVEVYSQLTQEEWHQVVG
SPIMLDSCLVDPTEESGALLVQKHQFQGRFLTPSSGEFKITSYPNLTDEHPPLPENP
QASAAFLPGYSRLVVEVARYSLAVSEFIKNLGSLSLGNHFAVGVDDILDAKVGGEI
IALGDLVGLRVLEV

Cpn_0705 793176 792334
CT671 hypothetical protein
FMELKKTAAESLYSAKTDNHTVYQNSPEPRDSRDVQVSPLEGKQTRQETKTSKGNTRTES
RKFADEEKRVDDEIAEVGSKKEEQESQEFCLAENAFQMSLIDTAAAGSAEAVVEVAPIA
VSSIDTQWIENTILSTVESMVISEINGEQLVELVLDASSVPEAFVGANLTLVQSGQDLS
VKFSGFVDATMAEADLVNTNPSQLSSLSVALKHQHLTKEFSVGNLVLQPKIEEVQT
PLHMIASITIRHREKQDQDNQKQKQDQEQDSYKKEEARL

Cpn_0706 793689 793180
CT670 hypothetical protein
YAVAKYPLEPVLAIKKDRVDRAEKVVEKRRLLTEIQEQLREKEAERDKVKNHYMQIKQ
LRDLQDQGTSSDAVLQKSYIKVAVQLSEEEKVKNQKEVLAASKELEKAEVNLAKRR
KEEKTRLHKEEMKEALKEEAEKEQDMLQQLPQLRQKVKKEGGS

Cpn_0707 795035 793704
yscN-Yop N (Flagellar-Type C/Pase)
VNMQLTDFDTIMQIADVNLTTCVQRIETEVQMLIKAVVPNVRVGEVCLVKRNGMEPL
VTEVWPTQIFAPLSPLELGGVSPSEVITPGLFLHIRAGNLLKVLNGLQEDTQVET
KPIAIPVQVPIFRAPDPLIRAKLRQILSTVRCIDGMLTVANQRIQIFAGAGVGS
GLLMIANNAEADVVNIALIESEREVREFIETDIEBIMKRVIVVTCDSQSRLRLN
AAVYVTAIEYFRQKQVTVTRARALREVLAAGEPIARALVYTVFVSTLPLRL
IERVGAIDKKTITATYTVIIVADINNEPVADEVKILDGIIIVLNAQAQYHPIADVLA
IIRKLLTAIVPEQRRIKAKPEVLAKYKANEMLIRIPEYPRGKREIDFAIDHIDKLNR
PIKQDIERTNYEEAAQKQKAFR

Cpn_0708 795738 795034

CT668 hypothetical protein
AFKTVKRFFCFMIDPVECF...DAEAQITQNDITFLASLKKDIDPFALGSAAPKD
TTLVQGFKNPMAMMDQDNHILIDPELOQALEGEELEQENINLKGRLWDFRSTFEEDSQT
AQFADEHFAQVGVITDLINEADLTAEHFOQDARKDEEGSVTRKIDWVSSGEVNLNR
ALLVFSRDGNRESLANFLKQYAVQQRATRAELFASIVGTSVSSVKTITMTTQGL

Cpn_0709 795203 795742
CT667 hypothetical protein
VPERMAKQENALPACIL...DAEAQITQNDITFLASLKKDIDPFALGSAAPKD
LHNMNIPARFELAL...DAEAQITQNDITFLASLKKDIDPFALGSAAPKD
QVIANKIESTRALQSVLLWHOTLVAKQAPLQ

Cpn_0710 795482 796210
CT666 hypothetical protein
RSRGEKSMATNKSCTAFDNKMLDGVCTYVKGVOYVLTETSTOQTVDLQTMFNQLFRM
QILSQVMESVSNLTAVNEMITMARAVKGS

Cpn_0711 795791 796486
CT665 hypothetical protein
TTINNVLGFINLYLGRYSMFNMENTAKEKNSQPLLEDQMDQHDRAQELKASVQDK
VHKLHALLREGSDMESFGQQSSLLAGYVALQKVLGRINRKM

Cpn_0712 799315 796781
FHA domain: homology to adenylate cyclase
MAVRLIVDEGFLSGVIFVLEDGWSIGRDSANDIPEDPKLGASQAIINKTDGSSYYIT
MLNDITIPVINGVAIQETTLQKNEEDILLGNSQYFLSDEFDPQDLVYDFDIPEENFSND
SGDLSDSNFGKDLPEQRTSETNHSKPKKEKLTQKQSGSDPITSGDQELADFLASAKAE
KNQPRAKVAKKLGESSNESLNPKEQNAKDSPKGEERTNKQNAIMEDNGASPRQDPQPK
SAEPLSKTARDETPLKENKPVEEKANKATPDSPKEKQDPPEEGSKKEGSKIETPLDSQ
KESEKDAEAFVQEEENLTEDNKEDSDSAADANDDTASDHAEADNKETPKVENEKSA
VLSPFHVQDLFRFDQTFPAETIDIAKGNISVDLTQPSRFLKVLAGANIGAEFFHLDGSK
TYLITDPTCTDIVNDLSVSHQAKITVNGDGGILLIEDLSKNGVIVEGRKIDKSTLS
SNQVALGTFLFLIDHHAADTVIASLSDDYSVLFGRQDAEALERQAEQEEEEKQKRA
TLFAGSFIITLVGGLAILFGICTASLFTHTKEVPLENIIDYEDLQVINQFPTVRYTFN
KTNSQLFLHGVKNSTDKSELYKVDLSFVKSVDVNDVDEDAVQWEMNILLSKRPEFGK
IDHSPPEPGFIITGVYKTEEQAAQLVDYLNHIFVLSLLENKVVVETOMLKAALHLLQ
GGFANIHVAFVNGEVLITGVNNDQAEKFRVAVQNSIGPGLVKNVFAVLLPAEGGIID
LNLRYPNRYRVGTGYSRYGEISINVVNGRILTRGVIDGHTVTSIQPNALIFLEGLKRYK
IDYKN

Cpn_0713 799817 799332
CT663 hypothetical protein
LDLKEEKHAFNEIVSIPQGTITTAALNTSMLEKLIKFNATYMGITSTLELDADGAYV
LPISSEVVKRAQONADNEIVLSASGLGAPSPADAKLYLQMMIGNLFGRETGGSALGLDS
EGNVVHVRFSGDITTYDFVRHVESFMNFSNDSLDGLGKG

Cpn_0714 801125 800091
hemA-Glutamyl tRNA Reductase
NRYVLMVLGVGISYREAAKERERAOYLOSFEKNFLAQRFGLKGGAFIPLLTCHRA
ELYVYSESPLEIAQALLSELTSGQIRPYRHGSLCFTHLFQVTSIGDLSIFGETIEQGV
KNRYLFGSKERELFPDLHLFQKALKEGKEYRSRIGFPHDQVITIESVQEILLSDYKSTY
TRAFVGVSDINRVAAYLYQHYHRTFCSRQOQVAFYRTLSTRETLFQRPQYVDFIFGS
SESASQPSDLSCEASIPKRVDFNVPRFLMKETPTGVYLLDIFISECVQKRLQCT
KEGNKAKLLLTCAAKKQWEIYEKKSSHITQRISSPRISPLSV

Cpn_0715 801636 803462
gyrB-DNA Gyrase Subunit B
KFNKISHMAAYTEASILSLASLDHILRLRAGMYIGRLGNSQKEDGIYTLFKEVVDNGIDE
FIMHGKSLKISADKQISIQDQGRGIPGLKIDCVSKINTGAKYTDQVPHFSVGLNGVG
LKVANLSEIFSVRSVRKKYKHLATFHRGVLOESQKSTKDPDGTFTVSPDPSIFPEFT
FNHDFKDKIRQYTYLHSGLEIRFNDVFIHSHGLKDLFADIEETPEPLYSPLFFQNEDEL
TIFSHLEGNTERYFSVNGQETLDGQHTLTAKEAIVKGVNFEKFTVSDNIREGIVGC
IAIKTASPIFESQTKNKGNTQIRSSLIKDVKEAIVQALRKDKVAPELLEKIKFNEKTR
KNQIFKQDLKSKQKVVHYIKPLRDKCFHYNDRLVGEASSIFLTGESASASILASRN
PLTOAVFSLRGKPMNVFSEETKMYKNDELFLYALALQITONEIQLRYNKNVILATADV
DGMHRLNLLITFFKLTLPLVNNHLLFLETPLFKRNTKTTLYYSEQEMQALQKFGK
KDSSLEITRFKGLGEISPEFAFIPGEIRLTPTVITSLSEISSILOPYMKNTKERRQF
IMDNLITDF

Cpn_0716 803466 804902
gyrA-DNA Gyrase Subunit A
FMRDVSSELFRTFMHYASYVILERAIPHILDLGLKPVORRLWTLFLMDDGKMKHVNANI
RTMALVHGDPAPIEALVVLANKGYLIDTQGNFGNPLTGDHAAARYIEARLSPLARETL
FNTDLIAFHDSYDGREKEDILPAKLPLVLLHGVGDIAVGMTTKIFPHNFAELLKAQIAI
LNDKFTVFPDFPSSGGLMDFSEYQDGLSTLRASIDIINDKTLVVKIQPOSTETTLIR
SIEIAAKRGTIKIDTIDQFSDTDPVPHIEIKLPKGRSAKMLDLFHETECOVILYSKPTVI
YENKPEVCSISEILKHTTALQGYLEKELLLLQEQGLTLDHMYHKTLEYIFIKHKLVSURE
VLAINKKISADDLQAVLHALEPWLHETLPTVQDTSQASLSTIKKILCFNEEACTKLI
LAIEKQAAIQKDLCKTKEVTVYKYLKGLLERHGLGERKPTQITNFKTAKTSILKQOQTLI

Cpn_0717 804968 805306
CT656 hypothetical protein
IRIKFIDTITVIRMEPRHIYIKRPETPKAPDVEKPGVPEYMTANTPTFEGPVKTLQDL
RRALIEQGAEEQKMYDNFIQSILISTFGLVHKMDPAQKASKMRMSVYKEQ

Cpn_0718 805300 805626
CT657 hypothetical protein
RAVMSFTYFALPVDRLMQERFLCSPKRWAPINCLPYLTLIADHDTTYLAKNLDKFLPL
VEGMEKTVLIVSLLKSLFLCGLSSRLLAQTKPELTLTLDLYAQNT

Cpn_0719 805877 806810
GlnR (translational synthase)
FDIPLVKKVIRIKMKTVTSTVCKENZRLDKYLTETVHFYQYQRIHLLQGLVINGQ
HTIPVATRLQGDHATIDIQKEELLELLPPIALDPVVEGMHILVINKIRMMVHPPAG
HPTFTVALLHIELEKKEEPPEEWRIQVHKLDPDTZILIITAKTRQAKVPSSELS
TPPLKEPILAVLCKKEETTLITHLSEHKKKEMTGGMEAVTHQVLAFFNKLKSLP
ALCHETPPIHLYNMHPLATITADIVVYIPMNHCTGGQKQDQJHAYVDFHTIPROF
GLVAPLELMKELLPEENNETTLINKNLESLKEK

Cpn_0720 806991 807206
CT658 hypothetical protein

106

CpN_0719 834892 833861

CT 164 hypothetical protein

CKVLFLKMSYSLRKKTKICVYIIIALGILSFRSIPQEVYDKIRSSFSVLHVKKFPKIKQ
APGSHLANLELNLVLRKVASLEELKLVYNNHTPLPFLPEITLPPYFKLVGKGVYR
YTHMSSCWVWVKTHGKKNPSVLSGNVGLVDVYVGEHQSRIILTDVGMKPSVAMR
QDQGWVTHGLRELIPQVEOISHAYLEKDYKIEISOLQELDGLIQEGEGNALLRGIL
PQVWALAPLILALIRHSPFSPKTKILIGLHGLDVPFGLVAMPYKVAERD
GATTFPILALIRKELMELALILIPKTHHIEFEDIPGLWG

CpN_0740 836054 834864

tyrB-Aromatic AA Amino transferase

SYMSFNHITPTSPDAILGQNVFADKRPKENVLIGVYEHQKRYGGLSCIRKAQTVI
LEEEQNKSYLPISGLQIFLDMRELVFGAVDPSAIVGFQSLGGTALHLGARLLSVKGS
QKVVYPEQWNNHIRIFSQECLVIRYPPYSEKQQLLEPLIAFLKEVEKNSVILLGK
CHNPFGVDFTEDMKELAILMKERELIPFDTAYQGFAGHIELDRKPIEIISEGNTVLV
AASSKNFALYGERVGVYFAVSTFTDELVIHSHFLEEKIRGEYSSPQWQVEIVSTILSN
PYLKEEWSQELNFIRESLQKTRFVQALRKVAGHTFDLLSQHGFAYPGFSKQVFL
REQHAYVITTAGGRMLNGITENKIDHVVQSFQAYEL

CpN_0741 838383 836185

greA-Transcription Elongation Factor

EYIFRLKTGDIIDYLEKQVLEEGQSANFLSLWEEYCFNDVVRGRELVEILEKVKSSSL
ASLFGKIVDTVPVLEKIPGKDKDRVLQIILDLQTSNSQMFDAIEYVNNKYSGEENF
NEALRVVGLRDRGDFQSLSRFDPLMMHKNFVHQGGVGEVGVSLFQKVLLEIFE
GIMSACISFETAFKSLTSLSGDHFLSRRFGDPDGFAPAKENPJEVVEILLRDLGPKTA
KEIKDELVDLVPADNRWQSAKTIKKGTRIISPDNPKEPVVLSAGCSHMGQLERK
LGLSLSAEKISLIYHIFRDLHSELKNIIRKSLVKALQDLVEEGNKLILQRELLSE
YLGIKDASIDKEYITSLSEDTSRLLNNMFIVALQKSLSLVRKYSFVQVQVFIQILLYT
TSPTMRDVFYKTKNDPSSVEVLKRLDLSAHQPMFPELFWFFKLGNHEDGLFDPED
KEVLRFLFESALNFMVQVASTRLLNMFVLRGQYLAVRQIEGASLFLKELLL
STKCPQSSDLNVLQSLAEVQPTLKKHKSNEEENVLWSTSEFSRMKAKQLSLVGKE
MYDNAKEIEDARSGLDRENSSEYFALEKRLAQEIRVLEEINRARIITKDLVTDKV
GVGCKVTLKGDAGEVEYITLGPWADDPSCILSLQSLKLAQNMGLKGLNDVILQGEYK
ISRIQSIWEHGA

CpN_0742 838442 838888

CT635 hypothetical protein

TKMMVIMNKSQAQIIDSIIQIILTYINIDFDPSPGSSSLSSDADYEYLITKQEKIQE
LDKRAQEIITQTMSEKQEVFANPNPDITSPWELALEKVRSSCDEYRKETENLINEITL
DIPHTKESRKPQKLSSTKKNKKNWIP

CpN_0743 838956 840362

"ngA-Ubiquitin Oxidoreductase, Alpha"

IFMKITVNRGLDLSLQSGSKESGFYKNDIPFVSDIRLPFQPLSLKLVQEGDAVCSGAP
IAEFKHPNTYITSHVSGVTAIRRNKRSLLDVIKTKPGTSTETTYDLOTLSRSLDS
EIEFENGLFALIKQRPFDIPAIPTQTPRDFVINLANRPFPSPEKHALFSSREGEYFV
FVAGVRAIAKLFGLRPHIVFRDLTLPEQLKTLIAHLTVSGPFPSPGSPSIHNSVAPIT
NEKEVFTLSFQDVLITGLFLKGRILHEQVTLALAGTALKSSLRVYVITTKGASFSLLIN
LNDISNDTLLISGDPITGRCLKKEEPLFGFRHDSISVLHNPTRKELFSFLRIGFNKPT
TKYLSGFFKKKRTYTNPDNMLHGETRPIIDTDYKVMMPRIIPVPLIKAVITKNFLA
NELQFLVEVCGEDFALPTLIDPSKTEMLITVKSLEIYAKESGILTPHQD

CpN_0744 841387 840389

hemB-Porphobilinogen Synthase

EMSSLLTSRRPRNRKTAARIDLAETHLSPKDLIAPFFVKYGNKIEEIPSLPGVFRWS
LDLLKEIERLCTYGLRAVLMFPIIPDDLKDAYGYSYSSNPNILCHSIHEIKNAPFHLCL
ISBTDALDPYTHGHGDFLAGEVLANDESIRIFGNIATLHAEMGADIVAPSDMDRIGYI
RSKLQDSGYSKTSIMSYSVYASCLYSPFDDALSSHVTSQDKQYQMNPKVLEALLESS
LDEEGGADIIMVKGAGLYLDVYIRIQTCLPLAAYQVSGEYAMILSAFQGGWLDKETLF
HESLAIKRGADMIISYSAFFILELLHQGFEE

CpN_0745 841903 841742

No robust homolog present in Genebank/EMBL as of 11/7/98

VSEDFDDWRASSLQSGSTTYNVAYDPKHTLAYGFCNQVSVKFKLPPKSKQKFL

CpN_0746 841939 843567

CT532 hypothetical protein

FSRCRPFSEVFMGLKEEFTCKQKQCLSHFVNTLSDVFLKKNLPEVVGALSKYSRS
VLGLRALLLKEFLSNEEDGVDCEADYDFDQKADDFQYRVLDNFGDSDVGLGGAHLA
MENVSILAAKVLEDARIGGSPLKSTRYVYDQKVRGEYLYYRDPILMTSAFADMLGTC
DFLFDYTSALIPQVRAIFKLYPKDSKTPASAYATSLRAKVLDCIRGLLPAITLNLGFF
GNGRFWQNLHKLQGHNLAEELRLQDESLETMLKVIYPSVRAEPHMHMMQAMQYRRAL
KEQLKGLAEQATFSEEMSSSPVQVVGDPDGIYKVAAGFLFPYSNRSLDLDIYCKMP
HEDLVQILESSVSARENRRKSPRGLECEVFGFDLADFGAYRDLQRHRLTOERQLLST
HHGVNFPVELLDTMEKSYREAMERNEYNEIVQEFPEEAQYVMPYANIRWFFHVNAR
ALQWICELRSPQGHQNYRTIATGLVREVVKFNPYELFFKFDVYSDTDLGRNLQEMRKE
PTT

CpN_0747 843949 844053

CT631 hypothetical protein

RTCMGCKGAEVQLSSRSLSOMKILSSSLFYKKFC

CpN_0748 844096 844121

ispA-Geranyl Transferase

CTLVLHALDTYRPSIESATEKALEGFGPIGHPIRSP/EYALQGGGKRLRPGVCMMAQGL
GLNHIDVMDALAVEFVHTLTIADDLPCMDNDDERRGPTVHKAFDEATALLAGYALIPA
AYSHLPLNAKLLKEQGCDFREIDIAINYIGDITDKHIGGSGVVGQYDDMFSSNRQGEHV
QSIMIKTKGSLFETACISGWLFGGGDQFAPITISQNNPGLKIDKDFSDQKDSQOI
GANTALLPGKAALELLARQNNCLLELLDRLSAGCLKNGSEFETIISLGSF

CpN_0749 845634 845006

glnD-UDP-GlcNAc 4-epimerase

VGYMPTLALIFEDHEDFLYELIKKAIYTWDLIDLMQMLNHFVGGHCTVEGGVTKN
LEKIEIAEYAVEYAYIVYDILGQVTEHRCYLRGNVINGRVCVGHCTETIKNSYLQ
HTRKAHFAYDIEVLEEVNGLVQVRCANRERDGHPIVYRTDCKSKIDTORRKLQAF
IKRWATKQNVVINKQILPHTRIRHQVFI

CpN_0750 846465 846407

ctcD/cpxR-HTH Train onal Regulatory Protein - Receiver

Domain
KITDFILRIHSYNLPCFHHIGDKIILFVTEDELSSSQKDLASQSDYQITLVSPVPTSF
ESVAIFCEYLLPEQIFSPGIPFEEDLIVLFDTFQEEAITKVLNQCATGYLLRPITAKVL
DAVIRAFRLQHEVLEHSIPDTMTFGDHTFVLMVLESEPGSVYLTPEAGILKLLINR
GHLCLRLNLAEIKNGTKEITARNVDVHTASRRKGLPGYSGKIVTITAGVGYLSSDDSTP
LQNHDNTAHNPNEE

CpN_0751 846408 846434 848734

CT641 hypothetical protein

MFCRILPGLIILLTCFSSGG/LYVLPQSHFISIGPKESRSWIEEKEFTDSVLHHLPSQ
HQHHLILCFQFGLLOKQKFSQAEEKIFSKVYDEAQDGFPLKEEILGSLRLNSFFLEKTD
VMETILCLLNQRCFNSPYHLFKALVYKQKLYREVIEQALYQWEKTRALAPLLNISIE
QLLTDPLDYISAHSLIEQKMFPEGVILNRRNLKHECEWNAKTYDRIAILRSYF
LELVESKSDIYFDYEMVLYFLKVIYIEQCPYAEELLPEELVSLIMEHVFILPKDKLY
PLIQLLEMMQKHYVHPNSLAVOJLVDRFSTHMEGATRECEALVSFSGLEELHQIITTF
EELLSNMVQKITEEAKQCVALLAILDPSISISEKLSSSDTLQNVISGDDQHTKLRY
LDLWEAIOQSYDIDRQQLVHHLVCAKDLWKKGKNDKALNLLQLVLRFTSYDIECESVVF
LFIKQAYKQALSSHAIRLLKLEKFISEANPTISVISEAKANFLADAEYLAHEHYDKC
YIISHMLTKVAPSPQSYRLACLMLKMKRYDEALEFLCLSPNDSINDYKTKALAFCKQ
HQSKDRAAS

CpN_0752 848595 850082

"recD-Exodeoxyribonuclease V, Alpha"

GWALHTFAPFLEDVHQVVISPLDIAFASKHISDDFEESFVFLAVSSALWRYGHPFLSL
EENRIRPSLGGISSETLYRGFHNLPKEARDKLFVVVSGRLVLSLYTIRSKLDKLSLLC
SATPNYFPPSIDSSIISEEQNFIFNKITQCGFSIVSGGPGTKTFLAAQLILSLVQPKP
LRIAIVYPTGKATSIHQIQLMKYNIIDDMQMOTVHHFLEQYAYRRYNSIDVLLVEGSM
VTFDILLYSVOTLQGYEKDKKLYTSLIILGDTNQLPPIGIGVGNPLQDILGYFHENTFF
LKTSHRAKTQVVDLTQSVLRGEMTSFSPSSISAIEVLKNRFVLSRQSEARLCVLT
MRHGPSWGLNLMIMHQRARSDDPLRIPIMVTSRYETWGLFNGDTGLLCLKQKHLFPQ
HEIPDSRALQSYVYVYMSVHKSQGSSEYDEVIIIPKGSSEVFGVSILYTAITRAKRYSV
WGPETLHKITKSNY

CpN_0753 851009 850161

No robust homolog present in Genebank/EMBL as of 11/7/98

IMATHALQOALLHRSWTPAIRASGNLFQKQSMHLNNVLFAGDIVGAIKNSTAIRHA
LGSSHYHAALQKTEGLGADGVNTAVAGAMLGQLLNGSMIFETDEETGELRRNCNE
ABGCMQKQRRSALTITGVARLASKTLGATFLHEDMVSVLGANRRYNSIDCVTSCLNL
VATGGLTSSISLYRILSTRPETISDPENKRRQSAEFAARSKARNAFIAWLDGVVDLV
CDALITSLFLPALGVHVLIMAILGLISCVINFPVQYAKIG

CpN_0754 851381 851040

rs20-S20 Ribosomal Protein

OFILNLKVLVLSGDIAPKPKKNVQRRPSAEKRILTAAQRELINHSFKSKVKTIVKK
FEASLKLDDTQATLSNLQSVYSVVDKAVKRGIFKDNKAARIKSKATLKVNRAS

CpN_0755 851579 852799

CT616 hypothetical protein

YKDLFMLLVKRWLHTCFKYIYFLPVVTLPLVCPYFPLSISQKIYGFVFTTISLW
FYALRRRNLKTAAVQLQTKIRKLTENNGLRQIRRESLKEHQESQAQIQSQKWN
LFLHQLGLVTKGEGQKLETLHLRTEENRCLKMQVDSLIQECGEKTEVOTLNRELAET
LAYOQALNDEYQATFSEQRNMLDKRQYITGKLENKQVDMYERINLLQLESIDIAENIPQ
ESNAVTGSLQSSSELKIAFAENIEAASSLTASRYLHTDTSVNHYSLECRQLFDSLR
EENLGLFVYARQSRQAVANALFTWTCYCAEDFLKFGSDIVISGGQWMDLHSSREE
CSGRILVTKSRGHLFFRYCLMALNKPCLYVGLVGLVPLHKEVLS

CpN_0756 852889 854676

rpoD-RNA Polymerase Sigma-66

ISYLPHTLSSKARNPLVLFQVRKLFMTQNSQATEVSSEESQKLEELVALAKEQGF
ITYEINEILPMSFDTPEQIDQVLIIFTGMDIQVNLQIDVERQKKEKKEGLEGLARTE
CTPDDPVRYLKMGTVPLLTREVEEISRIKRIEQAQIERILFRYSKAEIAHLY
ISGKERFDKIISEKEVEDKTHFLKLLPLITLKEEDTYLENLLSKQPDLSKQEAAL
NDSLEKSRIRTAQYLRCHFRNHTEDFGEVVFVAYDSFLHLEQDINDLKVRAENKFAA
AKLAARAKLYKREVAAGRTLEEFKDDVRMMKDSQEAKEVSNLRLCAVET
TNRGLSFLDLIQEGNMGMLKAVEKFEYRRGYKSTYATWIRQAVTRAADQARTIRIP
HMETINKVLRGAKLMMETKEPTPEELAEGLGLTPDRVREIYKIAQHPISLQAEVGE
SESSGDFLEDTAVESPAEATGYSMLKDKMKVEVLKTLDRDERFVILHRFGLDKQKPTLE
EVGSAPNVTREIRIQIEAKALRMHRPIRSKQLRAFLLDLEEEKTGTSKVLSKSK

CpN_0757 854709 855134

foiX-Dihydroneopterin Aldolase

PCIKNIALVIAIERYQLIISKFRMLFLGCSVEERHFQKQVLSVTSFVSYNEVPSACLSDK
LSDACCYLEVTSLSIEIANTKPYALIEHLANELFDSLVIISFGOKASKIDLEVEKERPPVP
NLNPIKFTISKELCPSPVLSA

CpN_0758 855104 856459

foiP/dhpS-Dihydropterolate Synthase

RAMSEPFVCLSLGSLNLRNFKMLQIARTLLGQAVLGRSSVILEALLPGSPPEWD
LPYFNSVLVGETTLLSRELLVTIKQIEKVVGRAEESPPWSPRTIDVDILLYGDESFCOD
TEITIPSLNLSRPFLLIALIASLCPYRRFCTQGSYHNTFTELGAHLHSPPGMIRRLS
PDTMLGPNVNTNDCMSDGMFLDPEKAVAQAEKLFTEGAADVDFGAQATNPVKQFLSV
DQEWERLEPVLRLKETNSNRKQYPIISLDTFYPIELIRAMDYPIQWINDVSGGSQMA
EVARDCLSLVNMHSGSLPVDPKNLSFSPICEGLDSWGEKQKMFSDVGLNANQVIFD
PGTGGKGAQAQSLATLYEIAKFKRLGCPILIGHSKRSFLSLFGNIDPKDRQWETVGLSIL
LQOQGVGLRVHNVAAHQKALSAVAAEACAPI

CpN_0759 856434 856997

foiA-Dihydrofolate Reductase

LLVKPVIQNFENPLVEMCKNRPVGRGIVACDPRQVITGLECKLPHVYFEDLOFFSETIQK
FPVIMGRKTWETLPPKYFVDRAVVFCHKRRKXVHGEIWTSLSEFLILLDSRPTFLIGG
GELYCLFLENQIVRDFFIHKKETANUTFFPLSLELTWTKTVIRDTQITTCYVENIHS
QNTKHSIL

CpN_0760 856994 857664

CT611 hypothetical protein

RHGPCLILIPKQKQVMTKITT/KTHIYVYDULYSLLEGLDKLNEKIVVTPKIKV
LGBAVVEKVKYDELIRQEAAYVYVYKVIYVYTKKWLITISAGIDENVEYVYVY
PQGLLVQNTLQWLRNRYHLEHGHIIIEHITPFLKRTMGLGAWNFFPIYVYVYK
DQFIRALKMYSNLLDLXSAVAVIARVDEQTHIAIEEAFKIPFIEGPTPLQUM:TLA
IAEDEDLYGLPLQMAWETPAIT

NLRKREQTLQVMEITLLPK[REDACTED]GKIPAPYPLGDKLAEDLSFHETIFRAIENKAVA
APIGIFPLKHLFPRGTHQDS[REDACTED]ENVLOWIRQVATECTPLSDGVISDRITAKGIPCAR
RTVAKYRAQLKILPANKPKKLEYIRGSGNHFRDRQF

$$u = 0.75, \quad v = 0.1, \quad w = 0.15$$

ELLYVSRFSEKONVGLNNMTRKISVYMLNLPORKKALEAAVAYIEKGFGAGSIMSLCRHS
 ADEHITKTGLSDLDLALIGHPKGVRIEIPFGSSGKTLATHIVANQKAAVVA
 IDAEHALDPYSYASLIGVNIIDMLISOPDGEDALSIAELLARSAGVDIVDSVAALVP
 SELECDIGDVHLDQARMSQAKRLTATNRSQTCVAVNIOIREKIVGSFNPETTTGG
 RALKFYSRILDRIRGISKSDNSIDTGLRKKVKAKNLAPPRAEFDILNVEIGSSA
 GCILDAVEYNIEIKKGSVFNQYEKKLGQGREVFREELKRNKLFEEIEKRIYDVAN
 TPSVHANKTPOEVAQTVEA

CPn_0763 860520 859972

ygfA-Formyltetrahydrofolate Cycloligase
 NFPMTPDKPIEKSLARKLFISIRRDLSSEERKHEASSAVAFVRSFSKESVVLVSFVSFNHEI
 LMQEANRILIQKCTLALPKIDQENLYPVLIPSIDLISVVHPKDPFSKQTPISDDKITHV
 LVPLGLAFDQOQYRLGYGHGFYDRWLQAHPYPSIRTIGIGYCEQKIDRLPQESHDIPLSQI
 YLC

CPn_0764 861819 860524

CT6748 hypothetical cellular protein
 GYKMDIKLFLCFLCSLIISMSPYIGTKGDYKEFLTLTGINIIDRNLSETICSEKELKK
 YTKVDFLAPQPYQVMRYMKNGRDNVSCFLTAYHTNGIYKOYLCENLRAGYGRYREWHN
 NIKIOIAEYIGGIDILHPSAESGGLVDFTQTTFYENDGELIAEAYIEKGLLEGYYRVYH
 GNINWKEPYHKGKVGQKFLFTYISGGLKLEQNTQYQSKRGLSLTRYSESDVVLWAEYEH
 IYVQVNTDQVDPDTEHYAIEGKQVYKQVYSTRALRYSGEYKGYKTFDNSQY
 IVQVNTDQVDPDTEHYAIEGKQVYKQVYSTRALRYSGEYKGYKTFDNSQY
 KSGLLTLYYVPEQIMATIEYDNDLLIKGEYFRGDRPHYSKIDRCQGVAVFFSSAGTITTK
 KIPYDQKPIPLN

CPn_0765 862415 861801

CT647 hypothetical protein
 TT IYKLLGRIMKKWISILILSFLSLLSILPVLAITINHVKISQRWSDLNSQILTQKVR
 DHEDQVIKHNAISKDRNNLSIESLNACQQLRPLSKERERLNKLNNSLSLAQSKVWER
 KRALEKSNHQLVWNCQMHNDAFVRLEQATEMDNEDIESLFLSFPNPVAPLVFFTCW
 KNTKPTPLGNEVWLTTHAEAISRWI

CPh 0766 863785 862394

1266 hyphothetical protein
 MKNKLPVYHIGLTKAEINTIKAIQLCKGWIVCHCEQIPGKTSWLSPKYKFAAFTTF
 KSSSDILVKKSSSSLNKNRNLINLNLAEASLAWESLIVQQLGKTFQDTPTLTL
 NIAKNTIKLKLSELSQAQIFPDKLSCRAADIFFLAEQSPKSLPKLPAVLLYGGSEVTCI
 NMAHIAVARSPNSNKKSCSDIDHATLYQITFTFPTPLPAIHVAQISPNLQKLEOK
 SSS
 ISILTKYALMATVYSLGSLKLLKSSASNNHAPPEECSVPSVPSVPSVPSVPSVPSV
 KSSSNPLLPITPTSEDTKFLLLAKSSSLEKSFYSFYSTYSPKDNPSLPYSALVE
 KGGQGPDIQFLKKISHPKLVHVSSELDORSEKLTFTLSS

CPn_0767 863878 864177

CT645 hypothetical protein
NIMLSYLLRTAINVYSFLILAYIFASWVPDCQSARWYQLVSKCVDPLNFFRRFVPRIG
DPSPFVGLLLCLGILPFVILRVLRFIILNIFHSPEWLLQYL

Pr-0768 864144 865163

hsl/nir/3-predicted oxidoreductase
 PHSFMAAPFIKNTLRSVVYPLAGFSDYPRCMSALYQGLMFCMKVVEILYAP
 RTSKLLDYNMRPIGAQLCGSNPETSGEAAKILESGFDLIDLCGCPDTKTKDGSG
 GGLTKPELIGRIDLKIINSVSIPTVKYISGWDHMINVEDTVIRIDAGASVAFVHR
 RYKQYVSGPSKQEIYISRAKAAAGKEPFGVNGDIPSPPEAAQMLTTCGDVYVARGTLGA
 WTGQIQDYLTVSGSEKIPITKRKAALHEHRLRDLSDVSETSKFSETSKLGGHYLIISA
 VTFPLRSSLLAKATSYQVQLVNDYEEDADDSSLETDFVRC

Pn_0769 867763 865121 .

opa-DNA Topoisomerase I-Fused to SV40 Domain
 IQGPAHRLMKKSLILVESPAKTIQLKLLGSEFVWSSIGHIVDPAKEFGIDVDHDF
 PQYQVLQPKQEVINIRHKLAAKECKEYVLSPPDREGEIAWHIANLPDPSLIQRVSN
 ITTQVNAIDTALKHVRPTIDMALVNAQQRRLDRIVGYKISLSRLQGRSGISAGRVGS
 AALKLVVDREKAIADPVEYVYNLRVLMQDPKTTITQFWAHLVAGKKWEKEIPGKTN
 VLLINSEAKHRHYALDELLKSSYITRVEAKRRVFPAPPITITSLQEQASHRFFPSAR
 MSIAQTLVYEGVDLSTSLGKLVVTRKQVADKAPPTATVYRIQOTFGFSPKEANV
 TTKTKMTQDAHEIRPTDINLPDOKLKNKSLDQFKVYHLIKRFVSAQITPAYTDLAV
 LITDTEIDLRAGSSGLKFKGLFAVEEKODDENDEEHLPLPHAQDALLKEVSEQYQ
 TDTPLPRTFEASLVKELEKSGIGRPSFTITMINKTSRFTTKENQRLRPTELGKIISQ
 LETNFPRIMDIDGTALMEDELEIDADNKPKFWKLQEFNTITFLPVVITAEKVAPIRIL
 INIECKSKHGKLVKIWSKNSYFYCCSEYPCEDYRSTSELANFKEDYADTDWDSPL
 GVMKVRHGRYGTFLGCEYPCGRTGILISKHGEEIQEIPICPAICGNKIKFKRRSR
 NIKFIYCESEYPCGVIGNSDIADITKYSGTETIPIYKTKTTPKKSAGKTTAAKTPSK
 YKNSKSVKSSKEKTKDPLFSLPDLAKMGNPVSSEATKKIWDYIKEHQLQAPENKKL
 VPPNNLNTIIGPNIDMPLSLKHLKSHLTLVNLNDESSASS

En_0770 868322 869131/

PF42 hypothetical protein
 PSTRNNVEKLEFVTLGSSPDGLITPNKQGLAGPEEEKVAFVLRNSNMLDAGPGETPASF
 LKREQFDLIPPEYVEVLVSNEDGLDWEAGCTWTILNIIEVTLQRKHRRKASPLWLGMSRD
 VLAHEAVIIAVRMKFIPEPVFEELVYQTSNWRFFGGLPFRSPQESYLLFFITLGLGI
 HYPACILIMLVVLMFLMRKLMAQSYVIRAMKKIIPKMLCVPPVLVLLRLTDKEIKMFA
 LITLPVLEIIVAPRKLENNWKKQIQGSEV

n_0771 870511 / 869144

101 RNA Polymerase Sigma 54
 P N G K E L L D E A L D M F Q K Q K L L K Y L P L R N Q S L Q M Q Q P L T E L S S Y V Q E I I D N P
 L S L S E E E E E F V P R P N I P T S Y L N N T P P Q E S L T R L L P Q L E A F S T A E E R I A H Q I
 L S L D D F L F L R N E D F A Q E L E K L I K W Y T I T L L S P E G I A S P S L Q S Y M K L R N S S
 A A G L V P D Q F L M T R E F A L M K K F S L S E L R H I L K A L K S I P W C P A A C T Q P M V S
 L M T L P G S G S G L N E V T A G L S I K L K E T F V E I L P K E E Q N L S I S A K W L I K

CPN_0772 872400 870469

uvrD DNA Helicase
KDLG IMTC ISELNEAGKAVTAPLNPLVVLACAGAGKTRVVTYRIHLHINOGIAPREILA
VFTFNKAARELKEP/INOCASTNEFDVPTVCTHFSFLGVLRRISNINLNRNNPTFYDS
DAEETLRIHLCHILFPLKAPDQAEYAKYKIPVFTFNTNINLNLNPLVYKSL
LEAKNIDFDDLLFPLKPLKAPDQAEYAKYKIPVFTFNTNINLNLNPLVYKSL
NVAFGVDDPDSYISWGAHNHILINENDYENAKVLLEENYKSGYNILNAANLKNNA
SRLEKELSVKGPEKIRFLFGSTDEEADFAAEILQHLVRNKLKRDICFYRNSNS
RTFEDALLRRIRYVIEIGLSFVKEIQDILAFRLIFISKSDIAFDRVTNLPKRGIGS
TIFALTOYAIAGQLPIKACQOOLLTDKVLKSCQOQEGLEQVLPFOIHAHYNTLSLR
DFIESVYFVLEIILEKADPTDKRKNSELVYKALSEBQNPKTHFLFDLLAKG
SDDDLNLTADRVNLTMHNGKLEFRFVFLVGELEQLPHANSLOGTYENTEERELCYV
GITRAQDLYLTLTAQVRSLVWVMPKPSRFLKEPKDYMIVR

CPn_0773 872485 873195

ung-Uracil DNA Glycosylase
FMQNATIDQLPSVWQEOELPLCWREQLKEEWSKPYMQQLLIFLKQEYKENTVYPEENCVFS
ALRSTPFDDQVRVILGDDPYPGKQAGLSFSVPVGGQRLPPSLINIFRELKTDLGIENHK
GCLQSWANQGLILLNVLSTVRPEPFSHAGKGWELFTDAIVTKLIQERTHIFVLWGAAA
RKXCELLNSKHQHOHAYLVTPHPSPLAHRGPFEGSHFSKINYLINKLNKPMTNWKKLP

CPN_0774 / 873183 873425

CT606.1 hypothetical protein
LEAPMNEGIHSVCFQKTPRLTAKSVVSMEMLLTTQQLPSAEGMPSVANLEADFLRAEALL
AEMREIRGCLQSLRTLVPSE

CPrn_0775 / 874040 873414

yggV family
ERFMKIV¹ASSHGKYIRETKTFLKRLGDFDIFSLSDFPDYKLPQEGGDSITANALTKGIH
AANHLGCWVIADDTMLRPVPA²NGLGPLSLANFAGVGAYDKDHRKKLLDLMSSESLVDRS
AYFCE³VLQVSPNGLIEFTKYICEGYISHQEGKSSSGFGYDPI⁴FKVYDYKQTFALSESDVK
NOVSRAKALQALAPHLQSLFEKHL⁵LRD

CPT/0776 874180 875487

[illegible]

CPr_0777 875586 877178

SEEDL-2 heat shock protein-60
 YRDELVNVVFKSGFGLSLAKRGVHLLKAVTPAFGPRGYNVVIKKGKAPIVLTNGKIRI
 KAGILQDAFESLIGLAKALLKVEQTGDGSSATLVIDALFTQGLKIGAGLDLQGEI
 KAGCILLSEVMYQQLQRAELQSPKDLVHMVAANHTDVTGTVAVISQADLPQVST
 KSKDGSIKTGRGLRKVSGYSPYFVTPEPTMDVVVEEALVLISHSLVSLSEELIRYLE
 ISEQNTPHYVIAEDFDQNVRLTILNLKRLGPCVCAKPGSRLRQVLEDLALTGT
 LITQGESEENCPISVDLNGRKVMITKTFPTFLEGGGDAEIIQARKQLCLTAARST
 ESECEOLEERLAPIGSPVOQVITADEQTERQFLQESALRAATKMGKGVPPGGV
 LFLRAAHAEVPAANSLSGQFTFETLLQAVRTPLKLVANQGRSSSEVTHILSHENPFR
 YNGMTDFEEDLDAGICDPLVITLSSLKAVSVSCLLLTSSFSISRTKT

Pn_0778 877400 878092

sa/ahpC-Thio-specific Antioxidant (TSA) Peroxidase
PVAQSDRVPGYEGGQRFESSLRVNNKRVEEVFMTLSLVGKEAPDFVAQAVVNGETCT
TKKDYLGKYYVFFPKDFTYVCPTLHAFQDALGEFHTRGAEVIGCSVDDIATHQQWL
TKKKQIGIGITYPLLSDEDKVISRSYHVLKPEEELSRFGVFLDKQGIIRHLVNDLP
GRSIEELRTLDALEFFETNGLVCNHWGERAMAPNEGLNLYFGTID

Pn_0779 878502 878095

T602 hypothetical protein
FDLIFQMKTVALFGAEKGSYDTAYFCRSLVDLHNYLGDVSSPGITLAIKTLSDYNV
YFVRVEEGCYDSYFFGLHLFNTQTTLKNI I AIGLPGVGNQHI IEASRSLCQKHNSLLL
FDHDLYLCTFNOF

Pn_0780 879241 878591

apQ/amiB-N-Acetyl muramoyl-L-Ala Amidase
 GNKIAVQSLRFMHAKLSFFILLSLFSGIDCSRLLHAAGRSPSLQGVLAETEDISAKLAS
 EVEIVMLSERLDEQDSKQKWTAAPKTPETLAQKIRLEESDQKALAKTLAVLTTSVKDLQT
 LQSKLQEIQKDHRLAQLDLRLVRSLLALVDSSPEGAYADFSDPVVPENITYIVREGDLSL
 AKTKYKLSLTKLLKINKLSDLSAIIYAGORLCLOPNKO

Pn_0781	879851	879198
---------	--------	--------

1-**Peptidoglycan-Associated Lipoprotein**
 NCYRSRRKTVPLGCFPSATDKENTMNIHSLWKLCTLLALLALPACSLSPNYGWEDSCN
 CHHTRARKPKSSFGFVPLYTEEDFNPNFTFGEYDSKEEKQYKSSQVAARNITFTATDSYT
 CEEENLALITLNLVHYMKKNPKATLYIEGHTDEPGAASYNILALGARRANAIEHLRKQGI
 ADRLSTISYKEHFKPLNSGHNLELAWQOONRTEFKIHAR

Pu_0782 881077 879773

IL16-polysaccharide transporter
 IL16MLRGLQVFFVFAGLVVAEELEVVRRCHITLPLEVQCGDTDKDKTKQKYLSGL
 IL16IKFDALIDGLQNTKAAKESSSLALSLRHVPOLSTVLLQSGKTPPTDCTITON
 IL16VQVQVLIHAADQVHALQVLPGLGAGKVFALSSGLQDQKLRQELWTLVDTKINLP
 IL16LITKRWKQVSNFFPYLVGVVYGVGVPLFGLSLEITGCKKLVKGLKQIMLITFS
 IL16KPLVAVKPLVAVKPLVAVKPLVAVKPLVAVKPLVAVKPLVAVKPLVAVKPLVAVK
 IL16VLPVLEIDVPLVAVKPLVAVKPLVAVKPLVAVKPLVAVKPLVAVKPLVAVKPLVAV
 IL16VITVLTITNKRKQWADIRKVFVSGNAEESLVLISLVTKTKNKILVGVKVRV
 IL16AEPQVQIKRL

71_9783	881038	381106
---------	--------	--------

50% hypothetical protein
IMKYLPR[IAITACIKKILLILVEACDCKKKKICDKAECEKIATITQKEIKVITCTWVWVWV

GYEFNELQHTFNCTLLLLLLDIDYHNGERLOFERTLLHLLAAGLSPDFPTFPKV
TFSSQHLRRPQLSGHFNWTFASDGLLIGLAGDITCLPITVALMARSLFLAYASSD
VSLQIKIDTADSGFKTTEGNEAVMTFIKYVEKDLLELLLSLEGAPTMFLORGESFV
RLPLETHQALQPCRLICLTGGEDILKYFQLPTELLLLCLPLNELNTDLSLMLANN
ETHSADQITLILFS

[illegible]

CPN_0794 897123 898004
No robust homolog present in Genbank/EMBL as of 11/7/98
KSKSHRSFLFLKSGGNGVSLYQWQNSQLKSLCYSTVAALFPMPSQSFADSLDLNL
GLDPSVFLGGDGAFFGVGYTITKAGSDTPEVEYQFFKYDWSKTTFTLSVETANQSGYANGIS
YDGTITVGTCSLGAQYNGKMSADGTLPLTGTITGSLTEARALISDQTVQEGFSYDA
SGQPKAVQWASGATVLTQLADISGGRSSRANYSVDSDDGTIVGSMSTITRKTATVKKWN
YVSTGCTGACGATGTVISGQDGTIVGAANTATVINGNOSHSGYMYKDNOMK

CPn_0795 898008 899195
No robust homolog present in Genbank/EMBL as of 11/7/98
GTGGAGSASTGVSDDGSLVVGQAATKDSVHAFQYYNGEMKDLGTGGTSSSTAKTVSPD
KGVHGRSPTAGDSWAFHMCFTDFSSNNVFLDNTYLTRENGRLNSIFNLNMLQQR
ASDIETFEGRSINLALGAGLYNALGNLPSNLAAQYFGIAYKIRPKYRLGVFLDNHSSF
VPNNFVSHNRLWMAFGIOWDSDALGSSVVCYFGYKQKATITREOLENTGSGESHF
BGNVIGBRYKSLGCHVYQWFLGFLGVHITIKETENAVQPVHYHDPIDYSTGVVYL
GIGSIALVDLSLHGTMRMGMEQFAHTDFSGSISAGNFVFEKLDVHTRAFAEMRVN
VLYPLQSLNLILRVNOPLQGVMSLSDRLYALGF

Cp_n_0796 899280 901340
no robust homolog present in Genbank/EMBL as of 11/7/98
SELYSSYLQCLNLSVRSNLPCLRSRSETTKKVRSHMKFMKVLTPWIKYKDLNVTAF
LLTALPGSFHATLDVADIEGEPHAAQTQVSDGGKIVIGMKVPDQPAITVGGQYIDGHLQ
PLEAVRPGCSVYPVIGTIPDGTIVGTNYAGMGSAVKVWKKGVELPMPTDLSVASA
VSADGRVIGTGNIRNLGASVAKWEDDVIQTQLPSLPMANNCVNGISDDGSIIGVTMDV
SRMTAVAPVIGQDLSEVITGLGGTTSVASAISTDGTIVIGSEADNSQTHAYAKYGVMSD
IGTLGGFYSLAHAVSSDGSVIVGVSTNSEHRYHAFYQAVDGMVLDTLGGPEYAGVGSQ
DGKIVTVGRAQSPVGDWHAFLPFQAPSAPVHGSGTQVTSQNPGRMVDNATISLSLKNQ
QQLQLRLISAKSVESVSGASPSTVKGAIKSGQSPAVQNDVQKGTFLSYRSQVHGMCQ
QQLLTGAFMDWKLASAPKCGFVAKVHGSDQDALVERAALPTDEQGGSSVLGPGGQGVQ
RYDFNLGELVTLPQFMGIQVLHISREGYSKKNRFPVPSYDVSAAFTSGHAFVPSLS
PKMSTALDQVERDLNSHIDFEGKSGVAMGNFLENSTVSVLRFPAASLAMYDVRRQQLV
TSLVMVNOOPLTGLTSLVSSQSYNSLF











CPrn_0797 901552 902694
No robust homology present in Genbank/EMBL as of 11/7/98
VLILTWIIVLTLMGLNMSKKIKVLVHLTLCTLFRGLVCAALANSIGYASTSQSPFPYQKSI
EDWKGYTFTDLELLSKGVSEAAHVSGNGSRIIVGASGAGQCGSVTVIWESHLIKHLGTLT
GASAEIGISKVGVVGVVMSDTRREGYTHAIVDRMDKGLTLTGATVSARGVSGDGIIS
VGVSAATGREDGYGVQVGVGVGKIKQKLKLPQLGVSEANAISDCTVIVGREGISRNHI
VAVYNNKNVAVSLGTLLGVVSAEAAISANGKVIIVGSTTNNGETHAFFMKDTHMDTLGTL
GGGFSVATGVSGADRAIVGVFSVKTIGEIHAFYYAEGEMEDTLTGCGEARVDFISSEGD
IGGSLTKDAGAERYALFLHIHK

[illegible]

CPN_0799 905001 903940
No robust homolog present in Genbank/EMBL as of 11/7/98
KREEDNAIAKIQILRLSLQSSLSLQVIGVGLSCGYCVITDKPEDHSSFAVKDWHMGK
TILRLNLNKAASAKASVAVTGVIGVGLKDTPTPTAVRWVWVGKELPTSSWVKSKATG
LSSGQGLTACTVNEVSGQAVVTKWNNEMLLPSTHAWGSKAGISDSCGVVSGAKDAW
LSSGQGLTACTVNEVSGQAVVTKWNNEMLLPSTHAWGSKAGISDSCGVVSGAKDAW
CTGGCGYCAIAKAVNSNGKVICRSTSEGVGHAFCHKNG/MSDLQTLGCGYSAAGKVSAT
CTGGYWRATNRTNCKHARYVYCGBMTDITGEYSWACAEAVISDGLVIGVQSGE

[illegible]

Q714_0001 0008707 006727
 0018 Extinctionless A15C Subunit II
 111TFTTFLQIIAIFAIKQVQVLAPELAAVRHQVKKQVILITTEITKPTFTTITAVANVMVH
 121TFLVLAHNKTLAAQLVQYRFFETFIHAYEVLFTYDYDYVPEAYIAQYDVTYIEKLLINDE
 131KKLKLATKRLTEKKRTKLTSSVQVLYGICGPNPNTMALVLELVQVQNPNTITQAVLK
 141MNYQAQLTPDQFAEAREFVLTITPAAYELALPLFLPIITLITDZSDPTMTIEKLEKVP
 151ATLTYLPEYHYVPEAIEKKAHITLITQELKEKRAFFPDPRTEKKLLENTITLITLMIKEIK
 161KFLKENVKRIETPAIAAIAITLITLITVDFEDELITIDEGYDTLITQTPAMVYKQKQKGL

mutL-DNA Mismatch R.          

CPN 0813 920843 921234

pepP-Aminopeptidase/P
 TLTLKKNHMSHDIRLRAFLRSEHNLDALIVKSESDLAYFLHDEAIGILLIGQGVNMF
 FVYRMDKDYLYIRQVPLITFDQVDAALSLVQKQRYQKIGFSDAAYVYHKFAQRQVLP
 CQWPLEPCEFTKIRISIISEEEIRMQEAAAGSAGDYDVLTLRLREGITEKVVRLRAFV
 AEAAGAPSPFPPIITAFESHPFSPHISITDPRLLKGDIVLIDIGVLLNGCYSDMTIRALTG
 THPHKLSESYVYQKRAMAKLCEGVLLWGDIAEAFAVRREHLDYFTHTGICGVGR
 HTHEYPCSPRGQSVLESGMTITVEPVVYFPGIGGIRIEDTLCIDKNKNFSLTARPVISE
 LVCL

CPn 0814 / 921996 923357

CT814.1 hypothetical protein
FFFLFKLSYNSPFLNPLMTMYQLLSIGYVSFVSIALLMMLCSYNNYTDLYRIISLSAESLS
GCTRAFPQAEISLLGCAGALVFPDLERLPDLRKELLFLGSSNDRPDCGCKGFSQLASSKE
CTIAALKERYVLLNSNSRGPFVSPGKQVTELWICFVSVDGVRVEVRLQGLHKLK
ISKPRDCEFLFNPNANKLDQWEIAGFVRDASFPVGGKQIKRIRIGVDFLLMHGGAEYADKA
TKERVDFSSDEENYSRYLAVGVLDWNGCWTGCEFGQASSRPLFEVKRIIDDKVIMIA
DLMNQVQVQCTITSLVKGPISEINEIVRIETIEFTGQSSWSKPIVLVGLQQLRSLPDQW
LRATKQVLEKLSRADQIQDVTYVQCTPLLVEFLKEKLDLGRFVLRGHMFNAQRTLVETISL
PLKQGFEPAAVSQEVSSNTRSAAGHPGATNRGGS

CPN 0815 923361 925622

gspD/pilQ-Gene. Secretion Protein D

MFVFNLSLHLVHLSGMLCCSSGVALTIAQMASLEHSGRGADDYEGMASFNANMREYSL
KGLKYEARKLKRASGTDEEALWKLIIRRGIEGVGYLTREELWAAEIREKGNLELYAL
NHHPETIYFNLVDYGTDEETKLIPOEIGAKTGLSKFVFNWFEETQGLTQISLRGIG
VFNWSMIKLLMYRRKGGCVAGVTSRRKLEALPTAYIGYVFLNSNVDATTHNHVLFKKF
INPETHTHVDVITAGRVFWSGAEVGEKKLYVNFVSGSIRQEVYVPLITQIDGEMISIL
NNAAFREEDTKDVQSEESLGRVFLVPOYQGRSLFSGTAALVQOALTLIRELEGNEINPTDK
TVFYNNKHSDDPQLAALLSQHVDVSGENKASVGAADGGCSQNASIQIDITVSSSKAD
GSKVYGNITADSKTGLTLMVVEVFLPRQIMLKKGLSDVKKVIRLEVLFAKHLAHOXS
LHLLRLRGEVCEKKGCSPPVSWAGGTGILEFLPKGSTGSSVPGVLYDALYLMQAEVRI
NAPSVPVFNITQKGVAVVMSISVNDKDKAQNRAQYVIMIKMLPVSDINGEEDKSG
TLETQIDTDTTGHHNDHDPDVTNRNNTKDIYADGETVYIIGLKRCSMDSHDGIFFLG
IDPICKGLFGMSSTDLSLTFMFFITPKILFNVEPQERKEEALLSSRPGEREYYQALA
ASEAAHAAHKKILMPEASGVSLGVSVEROEDYC

CPn_0816 925600 927102

sgp-Gen. Secretion Protein E

MSKPSNIAALSILQQLDILPYTLKKHCKLPIEESSEAITIAHATATSVIADQVKLLIK
KPVRFVKEESELQRLQQLNSRGRVNSDMLLTKMEKDTTISEEDLLETTDTIPVVR
LNLWLKAEIETERASDIHFPECDNRMIYDGLVHDHRSPPSHLRSLTFLKLVKLA
DIAEHLPDQGRKIHIGGQEGDVMRVSTVPVYIGERVVLLDKRNNVLDIAGLMKPTG
IKFLKDDTITAPGELLVITGPTSGSKTITVYVQLQELGPLNTIEDPEYKLPQIAQI
YAPVKPTGLTFARGLRHLRDDPDIIMLWGEIRPQETAEYIAQAKFGLHVLVSTLHTNDATIS
IAPLLDMGIESYVLSATLVGVAQRIIVRTICDCTAAVYATNQEKSGLASLQKDEMPLE
LQGVCGCHVCSFGYKRGQGIYFLRPNPTLPRSEVASNRPHYILRETAEQNGLPILEHGI
LAVSGGTTLAEVLVTRKRC

Pn_0817 927106 928287

spF-Gen. Sequence Protein F

GRMPRYRYTLDKPKERRRKGVLALHQAIEAREKLAQENIQVLDIRAVALRRMSIKSTEL
VFTKQLLLLLRSLGPLYGLSVLRDQYHEQKMLGLTFSMETLRSGGLSSQAAHPNI
DHFYCSGVAAGSAGVNLGECVNIIVLEAREKTKTKMVGALSPVCLLVFSAVMFLF
LEKPLSPKTEFEMEVKGLTKIVFGVSCLSAVRYFLFGFALATVGLIMRHRIPWKK
LGVLLPLPGTKTKIVGKVAANRPSVASAIKGGYTLFGLDGLCATPIDYRLKDMRD
VQVIVGGSLSGSLAQGVMPKLAIGMALGAEESGLADLVGYVAHYNEDTQKTLASI
SWQCPVILFISSLGIGLQVIMLAELPLTINSIQL

Pn_0818 928158 928682

redicted OMP [leader (16) peptide]
YTKNVGFDNVVSTRDSDFS~~WW~~PDRCOHVGNIDPTHKYQPNIIKCVLRGVGMKQRKQK
ITLIENM~~V~~VTILIGIIGGALFANMRGSIHKGKIVQSEONCAKVYDILMMEYATCGSSLK
IIAHKETVVEEAS~~W~~CKEGRKLLDKAWGEDILVOLNLDKGGDDLVIFSKRVQSSNKK

Pn_0819 929117 928956

T568 hypothetical protein
SLGYCLFLIWEKFHNNIGKANFHLKIITTDFLTDIYIVTIRDPIAYPLTGIC

Pn_0820 929042 929659

TS567 hypothermic protein
 ESLPGRCCCTGCTFFRSSETGSIRTEMPCMSIAMKKQKRGFVLMEILLSFTLIALLLGTLG
 WYRKIYTVQOKERLYNFYIEESRAYQLRTLFPMSLSSSYEEPGSLFSLIFDRGVYRD
 KLAGAVRASLIHHDKDORLELRAICNTKDQSYFETTORLLSHVTHVVLFSQRNPDEKPLPE
 TALTTIREPQKVPPTLLTYOFAVGK

En_0821 929637 - 930668

565 hypochlorite protein
 NPLNLRNGLNPGTITLTCLTCLAGLVAFDAANARKKCCACAAQTIERGENFFGKRSACA
 LTYQEKRRHGAALERTGKDKGVTFELAKVATPKKKQRYKRLQVPEGRDINNRNYLNA
 LSEPPFCEATATGAAWALILRLRRAYVDTGRLVPGPSEYAIALNLKQKLEPGAQGLV
 VPECLTLPQEAETPYKMLKQVGGGCTGLNLFVLKEGSKGKGLNLTIPMDLLEAVL
 HAYARETLLRLNLTWAAKVRHGLVQVQKAAALFLPFRTRDRLRLKMKOLLSSRY
 LKYLRLKMDYDILKLVGLVFLVDTFKALPCHCCTGKDKL

u_0222 930649 0000 000000

565 hypothetical protein
 LLIVLSTIKNISIGETMADEPKKEIKKEKGGDFDLKPKVKDLINPKVKWKKKFL
 PRACEALGKGLVAGLADFLHWGZFLFAGVVAHIVETRKMLCHLQYSILANGPK

[illegible]

CPN_0823 932424 231501
yscTdpAip-YopT Translocation T
FYALVRFSTKLT;INNGKSLGLCLPELFSNPKVADITFOHPHPAYVWSVFLLLLRLLP
IFVAFVFLGAKLFPSKPTIGISLWLAIFLPGSLADTITNYMNDNLFYVLVKEMIGI
VIGVFLFAVVAQAASGASFTITNOQIGLGBATSLISITESTSPHGLVHYFTITIFWLV
GHRVIGSLTTLTFLVPIHSSFPAEMHSLGLSTITMKMQLCLVLTLSAPAAHL
KQFLALITLTHHMAFLVLTITLALFAFMHSLPITLWVFIKIQDITFLAWFETFMIL
LITLITLITL

yscS/fliQ-YopS/fliQ Translocation Protein
IRTRAVLAFFATSFKSVLFEYSYQSLLLILIVSAPPIILASIVGIMVAIFQAATQIQEQT
FAFAVKLVVIEGTLMSGGWLSNMILREAGQIFONEYKWK

ERK-VOP Translocation R
YSKIVPTIMRSIFRFSLCFFTLVSQCFDASL~~Y~~ENSCPSRCQPTTPPPSNINPLVWVQPF
VAASSVPSYMPPLNADDVLPRDHLSDGSFSDTYPDITQTQAILIFLALSFPFLVMLLTSYL
KIIITLVLRNLAGVQKTPPSQVRLVNTIALILSYFVPTGVGVAIMRAKTEANTIPQSL
PTAEGAETVFNVALNGKPLRSLFIRLTPKAQIQSYNTISQKTPPEI~~DR~~HLTASDFVI
IPAFIMQGIKNAFIEGLVILYFFFVIDLVNTANVLVAM~~MM~~LSPLSISLPLKLLIVMVYD
GWTLLGLGIMISFK

yscL-Yop Translocation L
HDKNRKSSVSENVQPPRYIAIVMKFFSLIFKDDVSPNKKVLSPEAFSAFLDAKELLE
TKADSEAYVAETEQCAQIRQEAKDQGFKEGSEWSKQIAFLEETKNIIRVRREALVP
LAIASVRKIIKGKLELHPETIVSIISQALKELTQNKHIIISVNPKDLPLVEKSRPELKNI
VEYADSLILTAKPDTVPGGCIITETAGIINAQLQVLDLAKAFSTLIKAKNPVDEPSET
SSSTDSLSLNDNDQKLE

CT5260 hypothetical protein
GCLVTANTGPTGLDILMKHSEKDDSLRFLPKNLVESPHEEPIPLKLSFTMSWLPTIHPS
VITAMTKEPPEPIQGGLAWLPEPLVQEILPLPGISIAHPRCAPGFAYLLDMSKIR
PCGITEEFPLASSANAIIYTTGPVKIALINCLGLYSTIAKELKHILDKVIIVRVKNSLP
TEKFLTYTCQSHPMKHLETNPFLLSSWTTDAELQRFVHKQGLEFLGKALTKENASFLWVLT
RLDLVGRAYIVEQTLTKWYDHPVYFSKRLSEKPMVLVK

YSGY-Yop Translocation J
 YSGY-110MMVRASISFCLFFMLTLLCCTSCNSRSLIVHGLPGREANEIVLLVSKGVAAQK
 LPPQAAAAATAGAATEQHMMDIAPVSAQITTEALAILVQAGLPRMKGTSLLDPLFAQGLVPSEI
 DEKIRYQGLSEQSMASTIRKMDGVDAVSQISPTTENDNLPATSAVYIKHGRGLDNPMS
 VMISKIKRLIASAVPGLPMPVNSVSDRAAYSDITINGPQGLTFEEDIVSVGMILAKSS
 TPLPLIFPVYLLILFVTSQGLLWIKMTHTLIMTGGTGKFGNPTPYTKNALEAKKAKG
 LADKEKKEDASDQGESKNSDSDKSDKDADEPSEIIGA

robust homolog present in Genebank/EMBL as of 11/7/98

o-Robust homolog present in Genebank/EMBL as of 11/7/98
 SCSFLLPCFEVEAQTFPQVFSKVVVYKYKSSRILLIALLYNITLVGLIFTHKKYLGQK
 RVLTKIQNEEEFRATERPFSIGAGLVRVNNQSVLFPFEDMLVCPSPKDFPLSAF
 VTKLIYVWLESIPVWGAFSSGRLFAMWCIEDFPGSIFSRIYTHVGLGILGLGI
 MFILRIETLLTLPVGLISCKSSAA

Robust homolog present in Genebank/EMBL as of 11/7/98
 KRQNVLRKSESEGAFFETQNYPTIQGYQLVRIREHNLSVRAHFDLSLSLDASVHP

MKP-Lipoate Synthetase
 SPCAPRLTANTDQPRVKRLPERFKFKLQRLPQGSFAFHTATIKRSGMPTVCEEALCPN
 AECHSRKATATYLAGDVCTRSCGFCNGHSTPPALDPTEPERIALSAKELGLKHVVIT
 VARDLLEDGAGQGLVDITQKLEKLPQATTEVLASDFGKNVNSALHTLDSGGITVYNNHV
 YLPLNVRVHKAATATYRSMFLQEAANTLPDLIKSCGVMVLGEGEVEGVQTQLDLSI
 YRVITIGQYLRPSRKHLQVKSYYTPTETFDYYRVRGEAMCLFVYAGPFRVSSFNADMILA
 YOKASA

SVLA-Lipoamide Dehydrogenase
 LKDALEILITVSEMTQEPDVCVIGAGSPGCVYIAITAQSKLRLTALTEEDQAGGTCLNRG
 PSKALTAGANVSVVSIKHAEQGFIHVDGYTIDYIPAMAKRNTVVGIRQGLEGLRSNK
 LKDTGTGSLVSVSTVEYVIGDQTTIKANHLVLITDGLGPPRFVPPFSRLTSTGILEL
 LPKLALIGOGYIGCFPSLFTFLTGELVLTALDHLVANNKEVSGYVNTKPKTGSI
 LTKASIAISEQSONQVITVNDQVEYDGVVLVLAIGPQNTASIGLDNAGLNRDRDGV
 DETMRTNPNVYIAGIDTIGKWLAAHVASHOGVIAAKNLSGHGVHEWVAISPVITFTH
 AMVLGSLQEAQNLNPKLTFFPKPKAKVAGALGAGDGAFAIVSHEITQVILGAYV
 ASSLIGEMTLAIRNELTFLCTYVTHAFTLSLVWAGGALLATNHLHPPEKPS

55% hypothetical protein
IMFFAKETEMQRTWKKECEGYSIMHVDQCPYCSAFLQDPVVAIXXIFSSCHICFPEGASK
AEDFLAVSGEDWDAEAVDIDQNPQTQETNKQVITPEWTLQSWPLAALFLGIGLLAFNRLTL
ETDEHGLTWPKNRNAYFYVIGAAVAYRGYKLEL

IMVLEALAIFRODAHILKHKRKEIVVDFED:YTRILDEEAPESGYWLSTLKLODID
TFACDQDQDGEYGLILMAYFAVYDAIILHILHIDPFHESFWYAV:SHFFLDSDPLQAO
MYTTEIPHLTLTLEIEEVEFDWIRTHASEEPTVITNKTEKLSALYPTAKKFFL
EWAELEIDENGYEFIEHPSLQWYILVPEKALIDPITLEDIPKLELAHITGLENYHDI

brknQ-Amino Acid (Branched) Transport
 KKKKKNSSTHTDKKSLSTWSIGSIFAMFPGAGNIVFPLALGYHYNAHPWSAYFGMMTLA
 VCVPLGLGVSLMLFYSGDQKFFSIRPGMIFITLAILLIGLQIPRAIVASHFLIS
 LSEKSHAFPSLPLFASACVLYIFPSKLSRLIOWLSGFFVPMGLVTLVWIRSFMI
 THPMVQEFIPNAQWGLGFEIGTNTMDLAAFLVGLFSLVSLRQVAAEKKHPTTEEPIL
 SPQGISKKNRKALGWLGLFLLAAILGHTYLPVFLSAARHAGLVNLSKSGHLGRISALG
 PNSILAGVSLVYACLTETALGVADFLARVVSFKLVNSVASVICTLPTYLISLINF
 TSHLLPLGLVSLYPLALVACNIAAGNLFNFRYSFPLYLTLSLTILVKLVN

nth-Endonuclease III
LTMQKFLRLTNALFNPKNPSLEGWSSPFQLLIAILLSGNSTDKAVNSVTPQLFAKAPDA
QSILDPPGKLYQLIAPCGLGERKSAIYYQLSQILVRDFHGEPPNDMALLTQLPGVGRKT
ASVFLGIAYGKCTFPDVTHTLRLAQRWKISEKKSPSAAEKDLARFGEHNTPKLHLQLYI
YAGYCPCLHKKINDCPICISYLAKEANSTR

hNdf-Thiophene/Furan Oxidation Protein

MSLNIVPNSFLHMLKGLIGLSESSFNSIFMLKHDITAAIATPPGEGSIIVVRLSGPQAI
VIADRFSGSVASFASHTIKLGGVIFEEITLIDQALLMRSRSPSTGEDVVEFQCHGGFF
CNSQILDALILGARGPALPGEFSQRAFLNGGLDVLQVAAEQNLVAENTDAFIDIAITQTHFQ
FSGNSKKIQEITHTLIEALFEVLADPPEEQPDLLVPEQKTNLHVDRIEISFDEQG
LAGQSTSLILAGKPNVKGSSLLNKLQNRIVTHIPGTRDLEEWLQCKGRILRLDT
LAQORTLDEIDKEGIREALGAMAEADDLWVADITQPLEDKPLKFLKPSFLLMKADLT
PPPFDTLSLPOFAISAKTGEGLTQVKALIQWQKQEGATSKVFLVSRHHMILQEVAR
LCKEAQNLVLPQPEILALEREALHSGMLSGKEVSTILGEIFSKFCIGK

MSD-Phosphatidylserine Decarboxylase

720640 530141 591344
 T7000 hypothetical protein
 SERRRLKLTKEFFGIARQKSKQWRIMWLVLWALAAASLALVAKGYRFFVYFRYAV
 YIEVRVLSMLKELMAEQQLLILPKKRSYRQCLFMYRILRKMQRFESEKLLAEAK
 LGLRGSPFLFIEIYAKYAFRGAFAEKCAQAFASVPQDLFEEDAAQVASYALVRIGDLDAAC
 ILPEWISFSLSHQETFTVMGHTYITTSKRYKDAIDFYNRALMGVCPVEYTVLQYRIT
 SYAKAKGKFRKLLSVPYKEEALFVILGCLEQGRPGKALLIYQSSDLKQDGLDMLKY
 AMAAMDQRYVLAEPCWELARCTFAFKDYGKGLGYGFSCLRLRYKDGAEVRVYCNLLY
 PECILTACKALAWLCGVGYATLLGSEGGYMLAKYAKVELOHSCETDLELLSACEARGCNFDA
 YEIOSFLSSRDTSLOFKRQATSLRLTKKPIPNMHHVETDLELLA

```

ACCA-Translocase SecA
KRHHMLPGRKFFGSSQERILKKPKQKLVKYNVITDEMILPLSDDELRNKTAELKQRYQNG
LSDLSMLPEAYGVVKNVCRRLAGTPEVSGYHQRWDMVPYDVQILGAIAMHKGFTITEMQT
EGKTKLTVMVPLLYNALTGKPVLDVITVNDYLAQRCEGWYGSVALRWGLTLTGVLSGTLLE
PQKVTIQCVDVYVTSFEFGFDLHNSIATRLREQGVVSRALIDEVDSLTIDEARTPL
SGPGKEHNVPYVLEKKEVASVLYLQKELCSRIAREARGLDSFLVDLIDPKDKMVEY
GEPFRSLWLVSXKMPILNRVLRVREHPLDRAMIKDWVYHAEQNKPEESLRSELYEII
DEHNNDLQKIDPKGMQOOWVEYAGSGSTEFEVMDMGHEVALIENDETLPSPADIKNI
IAITSDETLPKARHGLQQRAALQLMEERDVIVRDDQIVITIDEHTGRPQGRFSEGLHQAI
RKEHVIDETIRKESQTLATVTLQNFVRLYEYKALCMGTGTAITESREFKEIYNLVVQVPTFKP
RIKHIDEFYMTEREKYHAIYNVEITHGKGNPLVDSVESEKSLRQNRIDEIT
NAKHIAQAEIITAGAGKLCAGTVATNMAGRGTDLKLDNEAVIVGGHVGITRHRSSR
QDQGLGRCARLGDGPDAKFFPLSFEDRLMLFASPKALTRHFRPFPEAGMSDPMFNR
LDTAQKRVCEGNYTIRKHTLEVNMKNQKQAIYAFRHDLVHASEVFDLAEILCHVSL
VMSLMDRQPGKWTLPNLEWITSFSPALNITELRQKDTDSIAEKIAELIQEFOVR
HHNMHVVEVLSKAGGEELDAICRDVVSVMYHIDEQWIRHVLVMDLLSEVGLRVTQK
PLFEKHESFLPESLIRDIRITIAHRLFRLELTVEPNPRVNVNIPVATSFHNNYNG
ELT/TTDSDSD

```

702 hypothetical protein (frame-shift with 0843)
YTPPTISRSPWSNTALKTISEPEYDGNQLLKQTSLTTIVDTLLNAPKDFPNSKNQKH
ECIAIRTLGHYAQELIAGNRKKEWILKYNQGVANSLTDEI

762 hypothetical protein (frame-shift with 0843)
 NKLITWRFKVMNYQYDTPDTDLIIFEDLLDNSEFAASLDKYQETGVVYEENTDQG
 LVLVGEVFKNTIHQEO

NCCTPase/ATP-binding protein

112

HLIIQCSROSYMGLSQILR...TFKQFPTAIHKOLLFFENRPFKOFDNRILRETAIVL
RHVQCSKAKVTFKPKYPTD...FYSYAKAHVYRFLPELNRNIPFQEQVGLLSLRLR
QYDSSLNTEF/PSAQHLLAWTRRGLALDLLYVLTGDCBKXCSQEFYINAPLLMSL
GNHTLEAFSPFYFYRANRLGFGTSRGTDMVLR/LATLMLSLRSEASTLAWSPKNLPSD
EABNVNSVFYVQGEHILPTFRGLPSVAGTSLVATHCSYSPENRLRQLYSTMLSLLVK
LHSHREMLKOLLPGTVLDSFETLSGGGLVFAESIAVRTHUNGAVSINL

CPB 0875 993363 994022

~~REF ID: A66087~~

CPN 0876 994123/ 995517

[illegible]

CPn_0877 / 995521 995982

RRRIMQLLSPAFYAGAPIPKKYTCQAGISPLTFVDVPGAQSLALIVEDPDVPKEIRS
DGLWTHWIVYNLSYITNLAEAGAEIFAVQGLNTSGKPVYEGPCVZKQHRFYFTLFALDV
VPDPEENKTDYDLSYFAEENHILFQAEILMCTYEKS

CPn 0878 / 996660 995992

SET Domain protein
GCMSTVTTFSSSIHISLNDWRDSQPSYSLDRASELLHFRFLPSLVFSNWKVEQOIETLC
HKSEKRRLLSPALKWLGKHKQDLLCPAPPVSCWINAHVGYGVFARDEIAPWTYIGEY
TGILRHROAIWMDENDYCFRYPMLPFLTYRPTIDSGKQGNVTRFINHSEQNAEIGVFS
PGLFHVIIIRTVAPIYADQICYHYGPLYKRRKKREEFIPEEE

CPN 0879 997463 996645

yycJ-metal dependent hydrolase
 YRLKRLKSGGIFPLASGSGNSAYLTGDSCKILIDLGVSKQVVTRELLSMNIDPEDIQ
 IFVTEHSDHSGGIFSPKVAKYNTPIVNCNLETARALCHLDSDPEFKIFSTGSGSFQDLE
 VQFVNVPHADVPAFIFHYREEKLGCTDLGWSTWITHELYDCDYLLIESNSPELVR
 QSORPDSTKIKRVSLKSGHISQEGGQLQKIKITPKLKLVLAAHLSSTECNTAELSTVSE
 SPSGTSIAPPAFLAQGTSFSGVSRQLLEVACPR

CPn 0880 999864 997444

FtsK-Cell Division Protein FtsK
PMIRKRSRHRLPLTPLAAKASLYLFAFCSGLSLWSFHRDPDCTQNWIGLLGWSFSS
LELYFFGAAAFPLILPYFLMWSLFLYRRTPRFFYKAAHLSLPFCASILLMSLSPVGLT
PALLDTRLPLKFLIGNPNVSVGGIIPFLFYEGQSGKHLKIGSVGALIGFVMLFVSL
YCGSIALLLKKKTQGVKAAFCFSFQTCFNKLKILNRRNYLPKPSVFNKPNKPSCTK
SSPSPRVSVETILDGSI5SLPQEEIPGSKSGSFFLTHPKCFRLTKFVPEQNKAECK
IALSSSTPVVRESKGKRAALPKLSLAVENDLPQYHLLSKNREARPE5LQAEERKA
LILPKOTLSPGIDADLNGICSGPTLAALEDFYHSGKVGKVKIKSLXENDIALKLOASRII
APLPGKAAGVIGITPPQPAOVNFRDLLEDYQKHTNRKLIKLPGLLKAGXNDLHADLATMP
HILIAAGTGGSGKSVCTINTVMSMITMLPSEILKLVILDPKVKLETSQYSLPHMLSPVITE
SREVNALLVWLKESMREYVILGLRNLNGAQFNSRNTKTEASYDREIEPTMPFVWGI
IDELSDDLLSDSGDIETPIILRLAQMAGVGHILILATORPSREVITDGLNKVPSRISF
SVSNVNSQIILDEEADNLIMGNDMLVLPVSGFTTQAGQYICDEGGINQIDLCPSRFP
TQYVPSIFLAFDDSDSDSGEKDPLFAQAKTLILQTGNASTTFLQRKLKIGYARAASLID
QLEAFRIGP5EGAKPROL1ENCLPE

CPn_0881 1005646 100620

No robust homolog present in Genebank/EMBL as of 11/7/98
NKKFAVHMPVPIDNSSRLQEVPESELDLEQHAESPTHQSAESSSLQLSLASSAISRRV
EQLSSVLGMENSDFFSLRDVPIFSAIYESSTHTPVPTPLVGVGINGSQSGYYDTQRES
LHLSQLGSRREVVVYNQGFMASLLNLCPRRPRRDPSPISLALLEWAEFFLEHPGS
TENIEFW

CPn_0882 1006169 100740.

[illegible]

CPn_0883 . 1008904 100757

dmpP₁NGRI-Phenoxymethylase/NADH ubiquinone oxidoreductase
 LYLEPFGKSGFIVTWLSCGLYFCIASLFGAICGLVAGCVILLGRKFLKVHPCKLKIND
 NEELTKTVESQGLTLLVLLSSGIVLIPSPCGGKATCKQCKVRVKVNADEPLETDPSTFKSR
 QLEEGWQKCKQKVQVHDSILIEERYLVLSASWEGYVINDNVATIKELVYANVQVQV
 FKPGYGLITITVQVSSKDKMKTAPVYGCWHEHFLDQGLQVQASAKNKKVLSLA
 LKQV
 DRPFLITLQAGSGSPFRSHLLDLLLHKSKRETLWLYGARSLEKNIVQYGEYENILERPQ
 NFHYHLVLELPELDFCAAGWCKDDPTKTHPLFAFNLGLQCLRIDNPEYLYVYCGPPLHN
 LKLLKLDQYVVERSSITLLDFDGS

Cfn_0884	1009668	1009900
----------	---------	---------

CT741 hypothetical protein
 GGCMLEIRIVTCFLFLLEENLPLFAEEEAQDKNTFFVQPAVMLAIAILFFYFLLWPFQKRR
 KAMEKKNDLAKDKVPMMLCTMDLRETFVLLHACKVEMKCALDFELKNDNKG

CPD 0885 1010650 100943

[illegible]

[illegible]

CPN_0894 1025823 1025888
amm-AMP Nucleosidase
PRNDKNGRNGKHYKCYERAPKHTKSSESRIADQMLERYSGSSVKQFCPYLLLTNFSYYIQT
FQDQVYVYEGFGSSAAAPKHLKTSI1DQFKLSPGSAALT1DLCFLPDLKAALMIGMC
CAKRSHYQCDYFVPVASRGESTSDAYFNEPALPVALFVVKATTEVEDLKKANYHIGIT
HTTNIRNENKFKRKLITVKAQSAEMECALTFAAGYRNNPLIGALLISDPLRKEGI
KTKSCNFIPTNTYTDHILTGQEVIENTLEKVMKRAASDHKKDQYRGLPHSDVEGADDT
MARGSETSDDSY

CpN_0895 1026973 1027557
efp-Elongation Factor P
EIDCFMVRVSTSEFVRGLRIEIDGOPYLILQNDVFKPGKGQAFNRIRKVNFLTRGVIERT
YKSGSEVETADIVERSMRLYTDQEGATFMDDETFEQEVVFWFKENIRQWLLIEDTYTL
KVDNDGWAVEPPIFMELSIETAPGVRGDTASGRVLKPAVNTGAKIMVPIFIDEGELV
LYDTRTGSGYSRFSVK

CPN-0896 1027574 1027822
 C7753 hypothetical protein
 EKYFFFRNRMEAKKIKELSKAQLLKKLREKSRVLDEKNRKAQWVAKLVAMPESIREIE
 KEERVLTPLFOAIAEKILEGV

CpN_0897 1028794 1027853
 (Glnsophyhdrolase)
 NFSLSDNSTVQDKMSNPRVQKPKGPDVRIIHISDVHFHVLVPNVPHVCFNKLKGLLRK
 FGLVHFQATITGCRVSPVSLGSAVSTGTFDPTAMGDFLAKHFVETLAKHSVYL
 LSLVLSLQAGTSLAQQTFTFNDQLQKQVSPFKITDHWMLILDCSLNGWGFANGV
 VHLAQTSIAETPLLSLSPEDVNIANKYPLLSDDHLLINTNLHQLMLKKYPKRVLYL
 HGHEHQAQVAVNCSLTPSYILNLSGSISLTPNSRFHVIDLYPEKYQVHTMLKNLLDFDAP
 I.EIANETWCKDL

CPn_0898 1030511 1028904
Mitochondrial HSP60 Chaperonin Homolog
TKRRLGQSVKILRLGLGVCMSQEQLSNYNADKKLFSGDKLQIVKGSYCPKQSLSPSTSF
KERGFVIAISQTSLENSYENLGVFAKAMWNKIKHESDGLTGILMLLAIQESYAALK
GISTHKLIASLKLQELKQLQALQEQAWPKIKDALKRVNITFSLIHMTIADHFYNFASVVG
PEGLISLTIKERENDTSMDFVGFQKIPAGVASTYFVSQTSASRLTRIAHPLILTRDKISM
ILSHLPLQIESSNQHLIFICEPIDIPDVLTLLVNNKGLQGVTVITPOLSTNQNELA
EDIALPTGTHICPGQASHVALEPMVTLGSLCSIESESQTLIGGLHIEPVLTAKTRQL
AEIEIRTSLELTKRLKIKSTRNLQSVAILPTDEENPEVTLALKILNESALSQRVPGVG
VALFVSFSTLGTDPKDDADNISLSLQKCAAPKLILATNSKLDGGDVAIKLSLGGTTS
LIGYSVRETIQDLAGGILSDLTSTTILAQALDAILVLSKLLILENQYISLT

[illegible][illegible][illegible]

CPN_0902 1034507 1035341
nlpD-Muramidase (invasin repeat family)
ADQDQNAKSEVNNRRNDIMTAIVANA LLVALVEFTSKRIGVKYDEGFPNFASFSSKVTQA
VVEGKVEIKPVAEVPETRIKATLEAAGIESKPVIVTTPPPVAVVEIEPTVIVAVPQ
PVRETVKEQAPYATVAVKKGDFLERIPANITVITAKLMQINDLTTLTKIQQIVKPTS
QDVINEETPTQYATANIENYIVGEGDCEFTIALRNHIPLDOLLPMNDLGEYKARRLKPD
QLRP

CNU 0903 10-6-24 193-417
 traw Cell Division Protein P33W
 CQGLGKMKWPFISGLALFESGLIMVTFITSAEVLDRGLECSTHKALIPVVTYLLILGIV
 ACYLMMHWRFDFIKFVSLAALALCFIFPLGLGICRRAWIAWIAVITLQIPSEFV
 YLIVFVITFETPGLQKQKHLFALFALFALFALFALFALFALFALFALFALFALFALF
 TVPGLYKWLFLDLFALGALALFMTFVRYRLNLYHLFELDIFKRGIFVYQAKLACGK
 ELIAGVGLLQKLTFLDEAQNITLALFVSGEPFLGMLVLVLYMCF777YALAKAS

116

CpN_0947 1117481 1134115
ydgB-like predicted rRNA methylase
LENGIFAIGPFMFAYRTLLTHVNVQVSHIEFKTFTVPGDTV DATCGNGNDSFLARLLQ
GEGRLVVYDIQKEALGNLLFETHLSEQERSVIEKQESHEHILEKDVLIHYNLGLVP
KGNKEITTLARTTEISLEYALNIVRPGDLITVVCYPGHPEGEKETHSVESLAQRHPKEW
CVSGFTVANRCRAPRLFIQRQSGESGVDK

CpN_0948 1117481 1134115
mesJ-PP-loop superfamily ATPase
KFFINLINDOGILKMGKAAFMHFFPVRSSVWLNRYSTFRICGPANYFKAIITEEARE
VIRFLHSINYPLIIGKSGCDFDRDGDGFLVYNAIYKQFLEDAIKAYSGLSFAALG
KATAYNGYSGLEFAAGISGCVGAIFMAGTNESDISSVRNVNITINSEGLCSYVEEL
ELSYRSSRFHQQEFILSATFQSKKQVSAHDSKSILOHRLMTQPYTOPSAGCIFRNPFG
TSAGKLDAAGLKLGAISPLHANFIINTGKATSDVEKQLIAIIQSTLKTQCIDLE
HEIRIIPYQPKIHSVSEK

CpN_0989 1139552 1139016
CT832 hypothetical protein
LRTSLAVKCVLLTIFWLVMTLSPEKFGSPISISKEFPQKQREIILQMLYALDMPAS
AEDSLVPLMSQTAVSQKHLVALAQTKSILEKQELDLIIGNALKNKSFDLDELVEKNV
LRLTLEHFYSPINKALITAEARILVKFYSSEACPFIOAILNDIPTDSSLNENSLSI

CpN_0990 1139880 1140440
infC-Initiation Factor 3
SVALNFKNRQIRAPKVLIGSAGEQLGILAKDALDLAREAGLDLVEVASNSEPVPCKI
MDYGVYRYGLTKKEKDSKAQHQVRIKEVKLKPINDENFTKLQARTFVEKGNKVKIT
CMFREGRELAYPEHGFVKVQKMSQLEDIGFVEAPKLAGSLICVAVGTVKTKKQKES
HAQDENQ

CpN_0991 1140394 1140612
r135-L35 Ribosomal Protein
KQKRNKSLMPKMTNKSVSARFKLTASGQLKTRTPGKRHLKSKSSQEKRLSKQPLVD
KQGVQMYKRMMLV

CpN_0992 1140622 1140996
r120-L20 Ribosomal Protein
GKLVMTVRATGSVASRRRRKILKAGFGWDRKGHRIQSRSSVMRAMAFNYMHRKDRKGD
FRSLNARLVASRIHSLYSRLINGKCANISLNRKMLSEIAIHNPGEFAEIANQAKKA
LEATV

CpN_0993 1140975 1142030
"phaS-Phenylalanyl tRNA Synthetase, Alpha"
KSPGSHSLGIRISMKEEBEIAVKQFHSLEDQVNSQALADLVRYLKGKKGIFRSFSEK
EALANVENYLLIEISCVSKSHADKALKESDFLIAGVQNVFSFLENQEDLYKSLDDEYSFT
KAYBEVKNKLEKVPYDLSDETEEHKEPECFNNLVVEVKRDSYELFYMLDEQDRFY
NDALVQIIVYKQKHLHETVNEGDPLTKTLWNSEEVKNIASSLVIVNDMPRLRYQALSH
LDIEAVVYKHNAMVALFYSRYEATMVFSKPKKHNIWYFNDFLLFLREAWKDLNNVIDSQ
BRKQTKLASALSIGIFESKLVFEEASRYLYFNIQTKLENANGKPLSPGQYLDAYEEL
HRLSEKYPNGPLFKAMDRLVHEHSRYPVDMILGLPSLEGTLLKHGKSIDIIISPSPVTH
SSILYANCNEEFLGLNAKHRSYVTLVNLQNRISRKERARSVIEEALEZEEHAFVYH
AFSEPEPEELLQNLLESIHGDIETPAFFSILQEEFHKPLASSFFLTKELEFVGSFLKE
KLTALKDIFFAKKLILFRNDKLLHLLSYLVFKLIERTNPNSIVVVSXGGLDYVSFI
AGFAFFSREAFWDEHSLKLLTNVLSPTLVARDRLVFVSHIELLSKFPVCLKNKRGQFSS
LKSFKKDDIEGWEFTGYLHELTEVSHKHL

CpN_0994 1142371 1144440
CT837 hypothetical protein
LFWHRRGGRKRSRRNFEOALENLEKLKEISLATSNDSYLNNPARFNQRKQSGSSVMEK
EALANVENYLLIEISCVSKSHADKALKESDFLIAGVQNVFSFLENQEDLYKSLDDEYSFT
KAYBEVKNKLEKVPYDLSDETEEHKEPECFNNLVVEVKRDSYELFYMLDEQDRFY
NDALVQIIVYKQKHLHETVNEGDPLTKTLWNSEEVKNIASSLVIVNDMPRLRYQALSH
LDIEAVVYKHNAMVALFYSRYEATMVFSKPKKHNIWYFNDFLLFLREAWKDLNNVIDSQ
BRKQTKLASALSIGIFESKLVFEEASRYLYFNIQTKLENANGKPLSPGQYLDAYEEL
HRLSEKYPNGPLFKAMDRLVHEHSRYPVDMILGLPSLEGTLLKHGKSIDIIISPSPVTH
SSILYANCNEEFLGLNAKHRSYVTLVNLQNRISRKERARSVIEEALEZEEHAFVYH
AFSEPEPEELLQNLLESIHGDIETPAFFSILQEEFHKPLASSFFLTKELEFVGSFLKE
KLTALKDIFFAKKLILFRNDKLLHLLSYLVFKLIERTNPNSIVVVSXGGLDYVSFI
AGFAFFSREAFWDEHSLKLLTNVLSPTLVARDRLVFVSHIELLSKFPVCLKNKRGQFSS
LKSFKKDDIEGWEFTGYLHELTEVSHKHL

CpN_0995 1145515 1144415
CT848 hypothetical protein
RMLFWKRRHLLTRFWALTSLLVLALIFYASIHSLHLTKGASTASGASVKSILVYLAQ
ISEKAEFLMPQVAVATTSTLFAQMKKREIILQASGLSKSLHPLLSGAVIMMVLVA
NFOWLHPICEKISITKENMDRGTDEKQKIPALYLLKQTVLYSSIEPKTLNNVFI
KDPKTYTMEKLAFTTSLPSGLNVTPQFANDSENLEKEFDMKEFPEIEFNFENPFS
KLFSAGNKNRLSEFFKAIPWNTAGLGLSTQVQRIISLLAIFYVYVLIPLACMAIILSA
YLCLRFSTPTVTLAYLPLGTNIFVFLKAGIVLASSVLPPLVMAFPLVLVFLTN
YAYAKLQ

CpN_0996 1146592 1145519
CT819 hypothetical protein
AMPILWKVLIFRYLKTAACTLLSICISISSLOEIVAYIAKDVVPTVLRIMAYOIPYL
LPFILPGSCFVSASFRLKLLSDNNKREITLTAEEQGDLDENQARKRYAFLYESYRQD
LASICRYOTCKEIANMAMTSPALLQTLQKKEIRIFIAVDHCAKSKFDNVIALKQNN
ISHVGIKSIIPDTTKDTVAKKDVFIISKLPDZLTSSSPSSQRYIETLDELLIPKITS
TLFAQKSYLKTTRTDYLPWKQVLQSKLHSHLETLRRAVIGFLCITLTAYAGMILGIHKPR
FRKSIALLYFIPILDILLIVGKNTKNLPLAFLVFPVQVSVVFAARAYRESRGYA

CpN_0997 1146699 1147664
mesJ-PP-loop superfamily ATPase
AYKMLGSDLLRDDKQDLDFASLDVKKRYLLALSGGSDSLFLFYLLKERGVSTAVHID
HQRWSTGAQEAKELEELCARQGVVFLVLTAAEQGDLDENQARKRYAFLYESYRQD
AGGIFLAHNDQAETVLKRLLESALTNLKAMARSYVEDVLLRLPLHIIKPSLKEAL
DAGIGSYLQDPINEDERYLARMRKLLFPWLEEVFGKIIITFPLTLGEEAELSEYLEKO
AQPFEEAATHQDQCELCPCDCLIQAFLLCKVMKKFNNAITAVIRHFLQMVYDHLRS
GCATLPMNRKIVIKGVVVID

CpN_0998 1147411 1150584
trfA-ATP-dependent zinc protease
LIGLVVFKM:KDKKKMDEPKMFTTFFVFFLLPQVVFVAFQNEIAGKARVQFHSQIEH
LVNLRILVPEDEHIALNDLV:FGGRFROVOTEGGLRYHILELIDXRHLDLDQETS
KQMTTUKEVNT:ILWF:YSGSPFPEQYAIISYRDEYGVVHPELVVTPATPOLNL
ILHLERYTILSRPHALPYV:DLVELICOKYLPVIGITGETLKRKLKDYQOVVSLTO
RTUFAAAYTLVQVLTNNRISGLSVSEGGERSFLLQVRLVREEMNKYHLVFAEDLN
QAQLEKLEHLEQVTVFNNQAEGLSRSLKQDPEVFIWFAQAKEMTAFKBNLELQFKA

PDQPRNLVLEKTFKSGEPS...FTFLPIILVLLFVYLVFGRQMRMGSGSAMSFGKS
PARMLLKQGNKVTADVAG...LELIEIVDFLKNFNKFTSLQGRIPKGVLLIGPFGTC
KTLIAKAVSGEADRPFPSIAGSDPFVEMVGVQASRIYDMFEQAKRNAPCIIFIDEIDATG
RHRGAGIGQGHDEREQTLNGLVEMDQFOTNEGYLMAATNRPDVLKALLRPGFRDRVR
VHMLPDIKGRFEILMVHAKRIKLDPTVDLMAVAFSTPGASGADLENLLNEAALLAARKDR
TAVTAVDVAEARDKVLKERRSLEMDAERKPTAYHESGHAVVGLCVQHGDPVQKVTII
PRGLSLGATHFLPEKNKLSYMKKELVDQAVLAGGAAEEIFLQDISSGAGQDISOATKL
VRGMCVEMWSPOLGNVTDERSDGLTGYGQVHEKSYSEETAKTIDTELRLMLDAAYQRA
LNTINSHKALELMTLMLKRETHAKDVKIMFNATCEKKNAPLPERMLFKKSSDCL
PMPKRETHALGATANT

CpN_0999 1152859 1150766
pnp-Polyribonucleotide Nucleotidyltransferase
QETFMNQTISINLTEGLIVFETGFIARQANGAVLVRSGETCVFASACAVDLDKVDL
PLRVDVQKFSSTGKTGGFIKREZRPSEKEILVSRILDRSLRPSFPYRLMDQVQVLSVY
WSYDQVLPDPLAICASAALAISDIPOSNIVAGVIGICINQWVINTPKTELASSTLDL
VLAGTENALMIHCHCDFTEEGVLDIEFGHKHVTICKRLQMQEVEGSKNLSAVYP
LPAEVLTAVKECAQDKTELFYIKDKKVAATAHEISENLEKLQREDDDLFSSFNKAA
CKTLKSDTRMALIRDRERADGRSLTVTRPITFVLPRTHTGSCSLFTRGETQTLAVCTL
GSEMAQRYEDLNGEGLSKYLYOYFFPFSVGEVGRISGPRRREIGHGKLAELKSHALP
DSATFPYITRIESNITESGSSSMASVCGGCLAMDGVPISISSIAGIAMLGLDQQAII
ILSDISLEDHLGDMDFKAGSGKGITAFQMDIKVEGITPAIMKKALSOAKQGCNDILNI
MNEALSAPKADLSQYAPRIETMOIKPTKIASVIGPGGQIRQIIEETGVQIVDNDLVVVS
ISASSAINKAKEILEGLVEVEGKTYRGRVTVAVGAFVEVLPGKEGLCHISECSR
QRIENISDVVKEGDLIDVKLLSINEKQGLKLSHKATLE

CpN_1000 1153193 1152891
rs15-S15 Ribosomal Protein
SAFAAILLRHMSLDKGTKEETKKFQHEKDTGSADVQIAILTEHIAELKEHLKRSKP
DQNSRLALLVQGRKLLLEYLNSTDTERYKNLITRLNLRK

CpN_1001 1153369 1153869
yfhC-cytosine deaminase
YYLELGEKLINMEKDIFFMQQAFKARKAYDQDEVPVGVVQKDDKIIARAHNSVEKLE
DATAFAIILIGSAAQDLNRLDITVLTCTLEPLCMAGAIQLARIPRIVWAAPDVRLG
AGGSVNIITEHPHTVSTCGGVCSSEAEHLMKKFFVEKREKSEK

CpN_1002 1153844 1154089
CT845 hypothetical protein
KSAERKVKIKVITLLDQYEDQKSLQKLGEEIVPNLTPEDLLQPMDFQLEGNPAFRFE
EGVLSGIGEVRAIALAALSQEN

CpN_1003 1154862 1154092
CT846 hypothetical protein
TNRKTIHPLWGPDRITACKASMRVIFPKHNNFNLKLLKPLSVILVTSIAPFFSY
IANKFFGICPCLLELALSVKGIQKHHPWQFLTYPLITADSLSNKQDSFEITQRLLRNV
LDLFFYKAIQHLIRKLGAFSVLVVISGQALIGAVLWGFMAHSHSSQFFPESIIICGV
LTVGIFLDEKPRITIGTPLSVSINKGFLVLTGYCCILLISFGAFLLLASMLATVAIL
FCKKEKIPNPTTSLRF

CpN_1004 1155418 1154879
CT847 hypothetical protein
HLSTIEELMSIQVSNMTTKADKVIIPDSTKVISDITINKQSAYFICISVRLSESTTEY
GKSLAVLEDNTIVQQORVKELINLPLDVLQKQDSDDEYKQNEIOAYQSSNQOIS
ANRQMIQOELSSAQQAQNAQKSVNSTTIESMQILOATSSMLSTLKELTIKANLTNSPSD

CpN_1005 1155957 1155415
CT848 hypothetical protein
NRKPVRLNMMIIDLPAKPLQAAINVPGTITGCPNTATADDIIAKFSKDSNPLIVTVY
VYVQSLVAQDNLISIAQELQANSACTYLNNOELQYQVSVIPKHLNDNSSSYLQNIQS
DNQAGASQAQIQNQISSLQAAQVIVSSNLNTNNNIQQSLQVQQAQITQFSQVSLIAN
I

CpN_1006 1156493 1155990
CT849 hypothetical protein
TKNVFFIMSTITLQTLPTVNTINSSRPPLPLNTPKIGAVLSFYELLQAIERQQTVL
TQSQQLNDNTNIQOQLNQETNIKYAIVSAGAKEITTRVQONQNSAQRSNIQDELVT
TRONGQILSHASTNINIQQOQSQDSFIIKTNSIGSTVNLQNLKPLG

CpN_1007 1156689 1156907
CT849.1 hypothetical protein
LWYKSLAGEKDVSGNECNDYPEVFKDDVSAYVLVTCQMSSEKTIQVEMTEBDPAVIS
YLLTKRADSLDES

CpN_1008 1156904 1158223
CT84850 hypothetical protein
VLNYSFQMLKPMYVLSKRLVWVQNLIKGLDVLKNSRFSVEMVIFISALLIFOCGLCA
SVKVSILVPLFLLLSFLAPFLILCFROKQYALLGVPTLVVAVYVGTETLYVSFNLSQL
QVDFLLAFGLFQGVWLAQEEEMVKGQELRLSDEDDAQRGAYEDLLTKSQKEFLDAR
AOGDLRELTCEQLLKAAYQKQYELTIDKLILADQKNQWLEDYALHNKYIELVSKNGDV
VFPWAEPVSQSGSERVDVSRWVSALQKEESLERLNEILVEKQRCSDYEHRCQELG
LLQNFATLERRCEELQNLNQKETOINELHQVLSKSEKVSVEPSAHAETSVCVEEKQYK
GLLYQLEQEFLEKSETLSLVRKKLFAVQEKYLLTKKKEELTKQDISFDDISMQGLLRI
EILEEVLHLEELVSRSLSL

CpN_1009 1159085 1158186
map-Methionine Aminopeptidase
YRLHRI/ILMKRNDPCWCCSRKWKQCHYPOPKMSPEALKQHYASQYNILLKTPEQKAK
IYNACQITARIIDELCKASQKQVTTNELDELQELHKKYDAIAAPHYGSPFPKICTS
LJIEVQITFINDIPKDGIDNNIDVSCIVDQYQDCSPMVIIGEVPEIKKKICQAALECL
NDEIATILPQIPIAEIGEAIEARADTYGQVVDQYFQVYIEFHENPYVPHYRNRSMIP
IAIKMIEITETIMMKKEAVDPPKNWEAKTCQNGFAQWEITIAITETGYEITLLND

CpN_1010 1159675 1159067
CT849 hypothetical protein
ZMLILLHLLSLFVLEPSTHIVVALLKNP:SRKYYRVILRECELFALIALILFVTCR
SFPQFLDELQYAKQILGQPLTPV:IKMMLAMPFAADDQCTP:IFPLAPFVITQPA
V:VALLGMEB:YGRLEI:PTAMI:IAWAF:CLP:LLC:CFDRL:GPN:GALLALER:GLAL
L:MKMLK:IL:IAFN:PYIG

CpN_1011 1160469 1159903
1154402

(TMS) hypothetical protein
FRLKKNYPIQGFPLPOTCILLASDGLTNILALH...VKQRM/LLRESFFAFI
AMFALYGLALGLKVLNTPVCAIEVVGIAVTLAGVRA...KESWI PYKFMSPSY
PCICPILPLMFGPSG

CpN_1012 1162220 1160421
yzh-B-ABC transporter permease
AIFGLITGKMKKKFIEYFVIVFGLFLEWMTSRHPTFFCFPPSGIAGTTLQSLPLL
AIFGLITGKMKKKFIEYFVIVFGLFLEWMTSRHPTFFCFPPSGIAGTTLQSLPLL
HIFSGGLKIAIGSAGFAAIGAGVWASGGLGILMESRRNYEMELAFAGLATLSITLSLF
QITLLIEKLIIFSLFVRKMSLKHKSVAKKALSVLALIPMLIPWKGNSKSPDKNLTSL
TLLDLPNPNHILYAGVAGYFQHGDLQKNTDSSAVPHVLEQVDMALYHALG
IKMTIKGHPIQIGRLIDSSQGLFYRSQDPYKFEELNGKVLGCLANRDLNRLLET
LNRKGVVPSSEKNSDLSIPMLNKLIFLYGAFNYIEGVKLTGLMPVKCFSLDTCOLP
TGPQLIVFTKGTKASEPEIVEAFQKALQESIIFSKDHPEDAFKLYAKETKSIKPNLYQE
YLQWEETFFLLAQSQDPLSKDLVKLLETIIKRYPELASEVAKFSLNDLYNPSLPEEQSV

CpN_1013 1162209 1163624
fumC-Fumarate Hydratase
RENSLNHRGNDIMRQKESLGIIVEPDEKLYGAQTMRSRNFWSGPELMPYEVIRALWVI
KKCAQAQNDLGLDLSKDCMIIAAAEIDEGFEEHPLKVMGTSGTQSNMNVNEVIA
NLAIRHHGVLGSKDPIHNDHVMKSSQSNVPTAMHIAAVISLKNKLIIPALDHMIRVL
DAKVEEFRRHDVKIGRTHLMDAVPMTLQGEFSGYSSQLRHCLIESIAFSLAHLYAIGATA
VGTGLNVPEGVFEKIHYLRKETDEPFIIPASNYFSALSDHDAVDAHSLATLACALTKI
ATDLSLFGSGPRCLGELFPFSGEGSSIMPGKVNPTQCEALQVCAQVGNNGVITIG
SRGNFELNVKPVIIYNFQSLSDLSGEMRASEFFVKGKLVNKARLQDNINNSLMLVTA
LAPVLYGDKCSKAALKAFHESISLKEACALGYLSEKEFDRLVVPENMVGHN

CpN_1014 1165456 1163732
yhm-Sulfate Transporter
ALASTLGYCIVKVPWAFKNFIPKLYTSIKEGYSFNTFKKDFQAGITVGLAFPAIAIAI
GVGVSPIQGLLASSIIGGLASAGSGNVLISGSPSAFISILYCLSAKYGAELFTVTLA
GVFLIAFGTLGLGTFIKYMPYFVVTGLTGLAIIIFSSQIKDFLGLMGANIPADFLPKW
IAYWHLTHWDSKSAFVGLFTLLIMTYFRNYKPRYGVMIATVATTLLVLLKIDIPITIG
SVYGTLPITPAIPKIPQSLITKILQMLDALTIAVLSGLETLLSAVADGNTGWRHQSNC
QLVAQGVANIIGTSLFSGIPVSGSLSRASIKSEATTPIAGIYHITLILLAPLTV
KIPLTCLAAVLILIAWNSEIHFIHLTAPKQDIVLLTVFILTVMITITAAVQVGMML
AALFPMKQMSDLSVITAKYFDKDSDFLSKAEPONTEIYEINGPFFGIDRLKLNLLN
DIKPKPKIFILCMTRVPTIDASAMHAEFFLECDRGTLTLLAGVYKKTPLADLKRYHLD
ELIGVDHIFSNISKALLFAQALTNLESKTSTRLHV

CpN_1015 1165550 1166893
hypothetical protein (possible IM protein)
KNNKQVFSFSTSVRVRKVDHEIILEVMTLKLQCALFLGYLAIVFHEIVRNVKSAIAL
AMGGLVWVCFSHIPMADHMLIVEEIDAMSOVIFLPSAMAVELIDAHKGFSVIVKPCR
IQGRTLLWALIGLSFLLSAALDNLTSTIIISILKRLVKAREDRLLGALCIIVIANAGG
AWHLEGDVTTMLWNNKITSWGIIRALFVPSLVCLVAGFCGQFFLRKRGSSTLIAKDOVE
LOEAPKSLWIIIFIGLSLMLVPMWACGLGLPFMGALLGLGLWLTSDWIHSPHIGDRY
HLRSPHILTKIDISSITFFIGILLAVNALSFANLLTDFSLMDKTPSRNVAIVIGLLSS
VLNVPVLAATMGMTPLDLDLWLKLIAYAGTGSGILITGSAGVAFMGLEKVDWLYF
KRLSWIALASYFGGLFSYFVLESNLFPI

CpN_1016 1167027 1168898
hypothetical protein
KREAFMKKGLGAIIVFGLLTSVAGFSKDLTKDNAYQDLNVEIHLISLKYAPLPWKELL
FGWDLSSQOTQARQLQVLEEKPTTNYCQKVLNRYVSLNDYHAGITPYRTESAYIPVVL
LSEGDHVFVDVQTSQGDIVLGEIDLEVDMGIREAIESLRFGRGSATDYSAVRSLSR
SAARGDVAVPSGIAMLKRRPSGLIRSTPVWRKRYTPEHIGDFSLVAPLIPKHPQLPTQSC
VLFRSGVNSQSSSSLSFSSYMPVYWEELRVQNKQFDSNHHIGSRNGLPTFGPILWEQ
DKGRYSYIFKAKDSQGNPHRIGSLRISYVWTDLEGLEEDHKSDPWEFGIIEIDHELE
TDALIIDOTHNPGGSVYLYLSSMLTDHPLDTPKRMHITQDEYSSALHMDQDLEEDVPT
DEGAVAVLGETMEGYCMMHAYASLQNFQSVLSVSWSGDINLSKPNLPGFAVQVRPHK
HOYTKPLFLLIDEDDFSCDGLADNCRATLIGKPTAGAGQVFOVTFPNRSGIKGL
SLTGSVAVRKDEFTENLGVAPHIDLGTSTRDQTSRFTDYEAVKITVLTSLNSNAKGS
EBOTSQPETEPIRVSYPTTTSAS

CpN_1017 1168997 1169935
Metalloprotease
IYIMRKLILCNRPFCSGGVVRAIQVVEALEKMGAPITYVKEIHNHRVNNALRAKGAIF
VEELVDVPEGERVYSAGHIPPVRAEAKARKLIDIDATCGVLTKVHSAKLYASKQYKI
ILIGHKHHVEVIGVGEVPEHITVVEKADVEALPFSSDTPLPYITOTTLSLDVQGEISS
ALLKRYPSIITLPSSICVATTNRQKALRSVLSRVNYVVGVDNNSNRLREVALLRRG
VPADLNNPEDITNIVNHSQDIAMTAGASTPEDVQACIRKLSLIPGLQVENDIFAVE
DVVPLPKELRCS

CpN_1018 1169895 1170629
No robust homolog present in Genebank/EMBL as of 11/7/98
RMSYFNQKNSVLRSLGLLAKFFSRLLYRVFFSFREGIYLFSSLYLYKYPFFYDLGKY
VYSLRHCPYAKLGRLPGLSLLKRGNYGETPWSVLAKICQAFDITSQDILYDGLGGLKV
CWFHSHVRCQVIGIDNPHIFRFSNNMRKLSGFAIPDTEEFKNVLSQASVYFYS
SPSRRLNLEIILKLSENAPGSVVISISFPLDSFGKCEFTTEKSCSVRFWPKTIAYKN
IRKGS

CpN_1019 1172146 1170638
hypothetical protein
IRPRHIMTSYQSTSTPPPEGEFDFVGDNATEAVAAEVQVALPGEQYAMLRATSEL
CFGLITGECALITQALPPKPELQEEFLVNGILMFSTSLPNLKPOQSQOTSLASHRNP
LAQQTLDNSTKASTETPSSFPFFSCAPEGDCS/DKTFTVQVTPKAEQGEQEAQSAQ
QAQGHVIRVYSDTIKEHIAKEKVSQSTKAEQTKHTQKQDATALGPMSTLYLHREVPQ
ALGKTKAKQKDEHRLDQCGQYEQEQEVEEKKTKPWTVEGQYTSNQVYFESYPTI
IPDFIVFALDEQQLVLAQKRVNTLDVRLCTELMPLMLKANDMTLRLEERLEPQ
AHEIAAZVIRQAKYARWIAATATLILGATAFNVGEISGDTIIPVQKIGGPFKDATAK
TFFPKDIPVFTLGLQTHAASKHLESEAVRAVAYEYKVEFMHODEVTRTIEEVKDNW
KMMHFLNLTQTHDAAGGLYQ

CpN_1020 1174694 1172150
hypothetical protein
GQNTMZWLAQAEVILNQDPIIPAPRQDESVYISYITVAIQFAQKLPKFEETK
GQDQKHEIPTHUKTEIATPRERILARGGHEFELINTDAQYTSWNWLFQKNTSEA
SKAIWQJEMFKPKTEKTKKALDKNLSKQYKQKPIITLIIHJHLKRLTVDGLYQGL

DGEQQLQSRREERSETY...EKKIETVQIKDLKDLFGQDQSNCKQKKSPPQD
DTKRNRIAKAAQAVPVP...FTLSYLLTKOGLSDFSYQCHDSVESTORELDA
LHEKRIETIKVGEKEKRERLQGLSDI...QWAPVYIGIGVAILSCGGIFAFAGFFAG
LISLVIKCLKELKFWDLKHLPIKNEELRRTITIQWVYLYTPVILSTIKLVENLFG
SPTEIGAIGQPAIESTMAALRCALIFSAEYIKLKGKLTQIQDIELKSFDRDDHYER
SQELLDNMESSFEALSRLINMYRELDQVYHLSLGR

CpN_1021 1174270 1173698
hypothetical protein
IRPRHIMTSYQSTSTPPPEGEFDFVGDNATEAVAAEVQVALPGEQYAMLRATSEL
CFGLITGECALITQALPPKPELQEEFLVNGILMFSTSLPNLKPOQSQOTSLASHRNP
LAQQTLDNSTKASTETPSSFPFFSCAPEGDCS/DKTFTVQVTPKAEQGEQEAQSAQ
QAQGHVIRVYSDTIKEHIAKEKVSQSTKAEQTKHTQKQDATALGPMSTLYLHREVPQ
ALGKTKAKQKDEHRLDQCGQYEQEQEVEEKKTKPWTVEGQYTSNQVYFESYPTI
IPDFIVFALDEQQLVLAQKRVNTLDVRLCTELMPLMLKANDMTLRLEERLEPQ
AHEIAAZVIRQAKYARWIAATATLILGATAFNVGEISGDTIIPVQKIGGPFKDATAK
TFFPKDIPVFTLGLQTHAASKHLESEAVRAVAYEYKVEFMHODEVTRTIEEVKDNW
KMMHFLNLTQTHDAAGGLYQ

CpN_1022 1175709 1174216
hypothetical protein
FSFFYALKLIQIMMPVPSAVSANITLKEDSSTVSTASGILKTATGEVLVSCALBGS
STDALISLALQCIILATQOQELLOSTWVHQLFLPPEVVELEIQVVDLLVQLEHAETTS
EPQETQTQSRSEQLPQOQSSKQSAISPRSLKLEISDSKQOQALQTPKDSAVRKHSAPS
PETQARASLSQASSSSQSLPPQESAPERTILLEQOQASSFPLSQSAEKQALTTSKS
HELKYKERDQDQOQEQHDKHDEQDEAESKQKQKRLGVEAVAEPEGNDIAALIFSD
QMRPPAEETSKKETTFFKPLSPMSVSRFIPSKNPLSVGSSIHGPTIKPVENVFLRFM
KLMAIRLQGAEEAEANELYMRVQKRTDDVDTLVTLISKINNEKDDWSENEEMKALLNRA
KEIGVTDREKYTWEEERKLLKENVMRKENMEKITQMERTDMQRHLQIESQCHQARSN
VLKLLKELMDTFIYINRFP

CpN_1023 1176008 1176331
No robust homolog present in Genebank/EMBL as of 11/7/98
GLDFLEIFIMKKVYLSIIFATYCAESLAVTVAVPLSEAPGIQVRVVGLOQFEEQ
GSVPYSFYYPDYGYVYPTGYTKNTQGESRECYTRFEDGTIFYECD

CpN_1024 1177317 1176334
xerD-Integrase/recombinase
IFFFWFSLSLKIAPIPLILKLSLASMTPSTQFHTTILEQFSLFLSDRGLCQOQSAIA
YRQDSSFSFTISAISSPQDISQNSVYIFAEELYRKEAETTLARRLIALKVFLLQDQO
VLEPYPPIIIEHPIKWLPSVLTPOEVDAALLAVPLQEKKNPHILAFRDTALHTLYSTGVR
VSELCDLGHVSDCIRVTKGSKTRVLPLGSRAREIDAYLCPPFDQOYKKNPHEDHL
FLSTRGHKLRSCVWRIRHNYAKQVTSKPSVPSHSLRHAFATHLLDNKDLRVIOEMLGHA
RIASTEYTHVAADSLIEKFLAHHPNRL

CpN_1025 1177266 1178879
pgi-Glucose-6-P Isomerase
GSAFVSEYREKTMERKFRIDCSTKILQELALNPLDLTAPGVLSAERIKKFSLLQGGFTF
SFATERLDDAILAALISLAEERGLHESMLAQOQVNVYIEGFPEMRPALHTATRAWT
DSFTGEAEDIAVRSRVAQRLKDLFTKVRQFTTIVQIGIGSELCPKALYRALRACYP
TDKHHVFTISNIDPDNGAEVLDITDCAKALVWVSKSGTTIETAVNEAFFADYPAKGLSF
YDHFIAVTCGSPMDDTGKLVLFVHLWESIGRFSSTSMGVGVLFGAYGEVFLQLLQO
KSNQOIALQPNARENPLMSALISINNRNFGYPTAEVIPPYSGGLEFFPAHLQOQCHES
KQKSAQQRGVRGFTSPVINGEPGNTQHSFQCLHQGTDIIPVEFIFGFEKSGQEDIS
FQSTTSQKLFANLQALACGSSENTNPNKDFSRKPSFVLSVSLNPNYSLLGELISY
ENKIVFQGLCWGINSFDQEGVSLGKALNRVLELEGADASNPEAASLLTFNFKFR

CpN_1026 1178961 1179137
itaA
CSFGPGKICEDRMFFIAVRSRGLDINGILAAKRGKQVVKSTAGAWIGSRGAVFYSLV

CpN_1027 1179172 1180755
No robust homolog present in Genebank/EMBL as of 11/7/98
NMPGVSVPPLSPVIVRVRSSSGSLDIOPHAVLKISILIFALVTILGIVLVVLSALG
ALPSLVLTVSGCIAVGLIGLILVRLILSTIRKVDAMGYDAVKEEQVLSRIELES
ENREIRDRNRVDEQCAHLSEENKOLRDPHILGHOMTERIASLEIENQALVAENTILLKDW
NASLSRDFRAYKOKFLGALPEKMDIACIEMQNLFLKPECIAWVSLPLETQRLFLYPK
GFOSLVNRFAFRSFFQTPKYEYNSRNEDNGKVAACARLKEFFSAVLGASCEYELG
ICERAVALKETPLPEAVYDTLVQEPNLLTAEHLKWEKCFYSYPLAPYSLVDYCKRFL
VOLFEELCKLFTTGSPEQALVRLFSYRNHIPAVLASGLPPPTGGSVYFLPKQEN
LWQSLIEVLAIRYLDKTFVRNSEWTSFEMFSYNEMCKEISGRIRFAEDYETRHSEEF
PSPSLGEGEGEEFLPPCSEEVSVLERPLDSDMMVWHPPVPGKPL

CpN_1028 1180995 1181999
mdhC-Malate Dehydrogenase
FFLKGVRAFKEVVRVAVTGGKGQIAYNFLALAHGDFGVDRGVLDRIYDVPGTALRS
GVHRELDDGAYPLLHRLVNTSLNDAFDGIDAAFLIGAVPRGPMERGDLKQNGQIFSL
QGAALNTAAKRAKIDFVGNPNVNTNIAKMHAPRLHKNFHAMLRDQNRHMSLAHRA
EYPLEEYSRVIVWGNHSAKQVDPDTQARIISGKPAAEVIGDQWLENILVHVNQNRGSAVI
EARGKSSAASASRALAEARSIFCPKSDWEFSSGSDHNPYGPIDELIFGFCRMLPSG
DYEIIPGLPEWPPIRNKIQLISDEIAQEKASVSSL

CpN_1029 1181987 1182844
No robust homolog present in Genebank/EMBL as of 11/7/98
RVFVISTMLWQVMSRQSPDELSQNAFKNIFNKQRCFIFCSLCCGPFVAFLLKLSRLA
PEISLSTGLGAPFAPSVICASAIIVQFLHLSKQOGETSKLCCAIKNTWSSLWLSLLS
MPFFIAMVAVTVAMLSLFGSLPWGKLPHVLTIFIPYLSATALLFLGSSFCCLFFCI
PVLHNOQSIDYRKLLECFRGNILRQFQGVVIALYPLALCSWLDLSYLMTHLVEIADII
TWSFLAQMFVLVPIALILTAVSFFNFSPSLRPLDSDMMVWHPPVPGKPL

CpN_1030 1183901 1182843
predicted D-amino acid dehydrogenase
FKVHMRIVAGAGYAGLVTHLLHLSQGTATIDLPDIPLGEASQMSGLLHAPTCK
KALKPPLAQDGINATHALITEASKALNVPISVSGILRPAIDEDQAOQFTEVEEFPKEV
BWEAKARCEISIDSMVIPNGLALFKGVTLNGLYICOLADAMKUTYQDELEIDL
ADIEEFYDHIIVTQANASILPELKDMFPMYQOLLEQWPKDLAMLSINAHKYVMA
NTQKNTLHATFEHNPPEETDPAIAYQIMPPVLLFLFGLKQAOQVLYNMRSSSKG
RLVPLGP(IREKIWPLAALSGKGLLYIGITDMLAQAVLRKATYIAKEPFTI

CpN_1031 1185567 1184099
argD-Arginine/Oxithione Antagonist
IRPRHIMTSYQSTSTPPPEGEFDFVGDNATEAVAAEVQVALPGEQYAMLRATSEL
CFGLITGECALITQALPPKPELQEEFLVNGILMFSTSLPNLKPOQSQOTSLASHRNP
LAQQTLDNSTKASTETPSSFPFFSCAPEGDCS/DKTFTVQVTPKAEQGEQEAQSAQ
QAQGHVIRVYSDTIKEHIAKEKVSQSTKAEQTKHTQKQDATALGPMSTLYLHREVPQ
ALGKTKAKQKDEHRLDQCGQYEQEQEVEEKKTKPWTVEGQYTSNQVYFESYPTI
IPDFIVFALDEQQLVLAQKRVNTLDVRLCTELMPLMLKANDMTLRLEERLEPQ
AHEIAAZVIRQAKYARWIAATATLILGATAFNVGEISGDTIIPVQKIGGPFKDATAK
TFFPKDIPVFTLGLQTHAASKHLESEAVRAVAYEYKVEFMHODEVTRTIEEVKDNW
KMMHFLNLTQTHDAAGGLYQ

GLIIVAVLWLTITIAVEIPFGAAKNTTPEIFTIE...SVSLVITSSVMQLAMLL
VYFSGNWTMLTITGVMLIPAYLASAAFPLKLSKSK...SIKAPLAMITGILGVVY
SLWLYAGGLYLFMALVLLALGIPFYIDAGKKKNAK...KKEIVOMTFIIGLLALTAI
FLFLTRIKI

Cpn_1032 1186153 1195566
CTJ73 hypothetical protein
IMAYTTPYTPARITGIGRQDDMPPOPFETFCYDGLLOAKIFENFHTVPTVSLPKEL
...
ALGLNFENAEPAKVN

Cpn_1033 1187656 1186187

CTJ72 hypothetical protein
NNKKDYSGEFLTTDTVDISIAFLPSEENFCYIKTILFFRVKKHYAFFYGEPMISFRLL
LSGLCALGISSYAEPTKETTGHYHYKARIQKKHPESIKESAPSETPHNNLSLSPVTNIF
CSHPWKDGISVSNLLTSVEKATNTISLDSFISLPQWPHYKALGQTQALEIPSQWQYFSP
STTWTLYDSPTAGQGVIVDSYTLIHYWQNGVDANQAAGTASSMNDYSNRENLAQLTFS
QTFPGDPLTLAGQYSLYADITGLYDNDQYSGFISYALSONASATYSLGSGTAYLQTFPN
SEIKVQLGQDSYNDGTNFSIYNLTKSKYNYFYASWTPKPSCGQGYVSLITSTRKVP
EQNSQVTVGSINAAQHIEKHLVLFGRINGATGTALPNRSYVLGLVSENPNRHSQDLIG
IGFATNKAKAISNNKLRYSVMEAFATIGFGPYISLTPDFQLYIHPALRPERRTSQ
VYGLRANLSL

Cpn_1034 1188589 1187732

Predicted OMP (CTJ71) [leader (18) peptide]
KTSWQYKYSYLSILVQKIRARYSMITWLFPTFLSCSSFYASCRYAEVRSIHEVAGDIL
YDEBNWLLDLDLTLQGGELSHSISWKAIOGLKQKQTPQEAWEAVPFWFIEIQM
GTVOPIESAIFLLIEIKQKQKTFVYTERPKATKDLTLKQLHMLNLSLEDAPOQAPL
PKNLLTVSGILFSGDYHKGGLDFLEICTPLAKIITYDNQENLVIRIGLDCQKYGIAY
POTYKAQELHPPYIFDNIAQVQYNSKLLSNEAALLRHQMH

Cpn_1035 1190081 1188570

aroE-Shikimate 5-Dehydrogenase
VVQLPMLVPIVHLQIWRFSYGVSVMLCATVSGPSFCEAKQILKSLHLVDIIELRD
LINEDDQELHTLITTAQNPILTFROHKEMSTALWIKLYSLAKLEPKMDIDVSLPKTA
QOTRKSHPKIKLILSYHTDQNEEDDAIYNELATPAEITYKIVLSPENSSEALNYIKKAR
LLKPSFTVLCMGTHLPSRVLSPISNAMYAGISAPQVAPQPKLEELSLYNSKLS
KSHYGLIGDQVDRSISHLNHLSSKLSINATYIKFPVTIGEVTFSSAIRDLPFGSLG
VMPKTAIFDHDVALDASAGCESINTLVNRQKILGYNTDGGVAKLLKQKINSVNNK
HATLVGAGCAKATAATLMOGANLHIFNRTLSAAALATGKGYPLGSSQILHEPFIKRNQI
IEKLPPLDPTWRFPPVMDINTKHPSPYLRAQKHGSLIHGYEMFIEQALLQFALW
...
PESCDSFNVYKFNMAKV

Cpn_1036 1191180 1189984

argB-Dehydroquininate Synthase
GVKPCSCRSCTIPTMLTGQVSEITITTPHVWKLISNFFQKLFSSISTAYPLVITDVS
VOHLLGLPILDHKMLGQVIVLTVPFPEGENKWTETFSIQYQVLDQNI SPKSSIIIGGG
GTMLDMTGFLAATYCRGLPLYLITPTITAMVDTSIGGKQNLINLRGKRLGTFYLPKEVM
MCPIFLSTLPREEWYHGAIAIKHGFADAYLWFLNHSKMLFSSQILHEPFIKRNQI
KAEVAGPYDRSLARKINFGHSAHAETLAGKTVNHGQAVSVGMETRIISLAGVYMK
TPOLIDQERLLKRFNLPSTLKDQGSIVPEHLNLSYSPENIITYLGYDKNLSQHELMK
IMHGLGRAAPFNGTYCASPMIEIYDILWSECHVMRHC

Cpn_1037 1192286 1191123

aroC-Chorismate Synthase
LHFSGRSRSLFLEELLRTSVSRSHYLVKVMKNSFGSLFSPTWGESHGSGISGVVIDGPA
GLELHESDFVAMKRRRPNPNTSSRKENDIVILSGVYKKTGTPLSLQILNTDVS
PYENRERLYRPHGSQYTYEKKFGIVDNPGGRSARETACRVAAGVAAKFLANQITFL
AYLSLGSGLTPLYHKLISPELILKHSIYSPYPLPNEKIQEILSLHDSDLSGQVIFSI
TSPIDHFLGELPGKVAHALSALMSIPAAGFEIGKFGASQMRGSQYDTPFVMEGENI
TLKSNCCGTGGITIGVPIEGRIAPKFTSSIKRCAVTTKKETTTRTPQGRHPCV
AIRAVPVVEAMINLVADLVYQRCSKL

Cpn_1038 1192750 1192199

aroL-Shikimate Kinase II
WKLELRNMTIILCGLPTSGKSLGKALAKFLNLPFYDLDLIVSNYSSALYSSABZYK
AYGDKFSECEARILETPEDALISLOGGTIMYEASRYAOTRGLVFLSVLPLPYER
LEKRLPERLKEAMKTLSEILTERIDRMKEIADYIFPVHDVHSSKSLSEAGSGLIT
LKS

Cpn_1039 1194011 1192665

aroA-Phosphoshikimate Vinyltransferase
VCFTMLTYKVPSSVYGNAPIPSSKSHLTRAILWASVAEGSKIINYLDSPTEAMICAC
KQMGASIKKFPQILEIVGNPLAIFPKYTLIDAGNSGIVLRFMTALACVFSKAITVTGSSO
LQRRMAPLQALRNFGASPHFSSDKSVLFTMSQPLRSAYSDEVCSDSQASALAVACS
LAEGPCGFTIEPKERPWFDFLSWLKELHLPYSCSDTTFYFGSSHPQCFSYHVTGDFS
SAAFIAAALLSKSLQIPRLINLDLIDQCDKIFPSLMQNLGASTQYDNEILLVFPSSFS
QGISIDMCCIDALPILTVLCCFADSPSHLVNARSADKESDRILATIEQLKMGACIQPT
HGLCLVNPFLYDVLDSHDDHRTAMALIAALYASGDSRIHNTACVRKTFNPNVQTLNI
MEARIEECHDNYSMMSTHKKRVFARESFG

Cpn_1040 1194876 1194073

No robust homolog present in Genbank/EMBL as of 11/7/98
RPSQSLFRLTWSPSSPREHTVCAAPLLYPRRSPDYLFSPGTGCMSTTTVKHFIHTASR
WEPVLYIVAGNYWHAQWINTLSFLENSGAKKISASEHPTVEVKBEVLKHAEEFRHGHY
KTQIGRIGETSLPDYTSKNLGGLLTKYVHLHLDLRTCRVLENYSLSGQTLKTAAYILV
TYAIELRAGELYPLYHDLKEAQSKITVKSIIIEEOQHLOEMRELKDLPHGEELGYAC
QPEGLCIAFVERLEQMIIDPDSSTTKF

Cpn_1041 1196304 1194726

bioA-Adenosylmethionine-8-Amino-7-Oxononanoate
Aminotransferase
DPRRIFLHIEKTSRACQVFLVMPDFSEPSISRYKCHLFNCRNALTKPOYLRYGCK
WYRIGFALANPQAYNKDQCKRNKHNKHLCTPMESCMKQSGNCCIWHPET
QALDCTTKIVRIBAYLYAEKTRVLDASOMNLYLGHCHYITKKLCEQAKLEHV
IFANFTHIEALEVAKLAPLPECLERFFSDNCTEIEIAMKIAVQYNYIKAKSHFV
GCHAYRITPAMKIACTRTVYVPHDLEPSTNAPYKKELAIAQATYFSESMI
AAFIYHIAAAYAMMYNPELKEILKAKHNYLADDEITIPRTGCLPASEFTDI
PDIHICLSEHLYYHIALVTVTKETHDAPVQDHMYALLHITPNNPLCCSAALASL
DPTIEPRTAQMTERKIQEPFAEIMWOREVIAVTLALDYPAETATYFSQYRDHLN

RFFLERGVLLRPLGNTLV...QDELRITVSHLODALCLQPY

Cpn_1042 1196629 1195934

bioD-dethiobiotin synthetase
NRSPFTYFRANFMQRIIVGIDTGKTVGAILARALNAEYWKPIQAGNLENSDSNV
HELSCAYCHPEAYRLHKLPHKAAQIDNVSTEESHICAPKTTSNLIIETSGGFLSPCT
KRLQDDVFSSWSCSWILVSQAYLGSINHTCLTVEAMRSRLNLCGMVNYGPEDEEHLT
QETKLPITGLTAKEKEITKTIISCYACQKWEVMTSNHOGIQGVSGTSPSLNH

Cpn_1043 1196771 1196773

bioF-2-Oxononanoate Synthase
PMLCQOFLIEALARRKSKHTYRSLSNLSHIDFTSNLDYFAGSSPELRKEYITKLHAIES
LGALGRLTGHSQLCQRIEEQAAHYHNSFCLEIFNTGYTANGLLYALATQDRIHLDL
YIHASTYDGRLSKAQSFPMHADLNHLEKRLASSHLGRTFVCVESVYSLHGSVAPLQAI
SELCEYSAYLIVDEAHAVGVCDQGBGLVSALGLQDKVLATVYFGKALGTHGAAIAGS
SILKMDYINFCRPFYITTAQPHALTAIELAYEHNRQAFNRQREHLSALIHFRKAQNLG
LOKNDYITPQISICVSGSHRAQAALQIONSGYDVRPIVSPFVKOREELLRICLHAFN
TKNEIDHLHTLEQIFLCHVSSL

Cpn_1044 1198700 1197699

bioB-Biotin Synthase
AKHREETYSWSELDIREIYHTPVFELIHKANAILRSNLSHSELTQCYLISIKTGCCVED
MAYCAQSSRYHTVTEPMKIVDVVERAKAVELGATRVCLGAARNAKDRYFDRVLV
CIVKSITLDLGAEVCCALGMLSEEQAKLIDAGLYANNHNDSSPEFYETIITRYSYEDRLN
TLDVWNSGLITSCCGVIGMGSEEDIRKAFLLVTEARDHIESVPVNLWPIIDGTPLQDQ
PPISFWEVLRITIAVAVVFRSMVRLAAGRAFLVEEQTLCLFLAGANSIFYGGMGLS
NDIDEDAEIKILGLIPRPSFGIERGNPCYANNS

Cpn_1045 1199602 1198901

conserved hypothetical bacterial membrane protein
GTLPMNTSHRKTIVFSYLSSTFTLLLVLSNLVSSKLIPTFNFPIIPGLILYPLFTLI
SDVNEIFGPKARVMIFSAFIANLASSIVQIFMFPVPASPEMTAWHCLFDLSPLRL
ASLLAFIVSQDLIVLYTFKNRTPNSSLWLSRSGSTWISQIPDTFIVDTCLYFGMGLS
FPOTLIMFYSYIYKITFCVLTPLFYLAVENTIRKFLGMPSTKIANTVPLINQP

Cpn_1046 1200675 1199590

Trypophan Hydroxylase
VHYCEARTLPQYILKIALKRLQSLSLFFQNSQSLQRAYSTPYSYRIILQKENKEQKALA
RHCKTSLIEFFKNLLFVHLLSLSKNQREGCTDMAVVSTPFFNRNLWYRLSSRFLSKWS
YCPHFDLVLEAFGLSDFLDHQAVIKFELETHFSYYPVSGFVAPHQVLSLQDRYFI
AKHRLKLDNDFSLTPDLIHDLGHVPLHLHPSEFFINMGRILFTKVIKVALPSKKQ
RIETLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQELGHAFIDNVRVPLEL
DQIRLPNTSTPOETLFSIRHDELVELTSKLEWMLDQGLLEIPLYNQEKYLSGFEVL
C

Cpn_1047 1200537 1201343

dapB-Dihydrodipicolinate Reductase
FGSRNMGSHVGVIGSGRTGKIVIAELOSSEYTLGPGFSRSSALTFOVIAHNDVLV
DFSHPLLEKVAHLLISPKPLIIGTGFGPKCKEADHSELEETHIVPVCNPNASLJG
YHKLRLVLLSQCNPQDIRIRETHRYKDKSLSGTAQDLDLDTIQOVQEDWEEYEVQ
RDGSKTIEVQSRVGDIPGEHEVAFISSGEQILVRHTVFSRVNFGRLSILDLWKLTLN
PQGLSYSGDTLELVRNHECHLKKTTDH

Cpn_1048 1201588 1202604

asd-Aspartate Dehydrogenase
DGERKQIRIAVLGVTLVGKFLVALLHKWYDVIAEVVASNSKYQSGYDAGIWOEPIG
FPMEMVRLPIRKIEVSDIVVSLPSSAESMEAYCLSGQKVVFSNASTYRMSHVSPII
IPEVNSDNHQLLEQFPYKGIITSPNCSCVGTALAPLAKFLSDHVIIVTLOSAGGAY
PGVPSLDLLANTVPHVGEKEILRETYKLGSSQWPKLCKLSVTVRVVPAYGHTLSLH
VTFKMDVLDLILYSQEKNEKFPNTQYLDNPMSPQARKHLSHDDMRVHLGPITYGGDF
RTIKNLDIHNLRGAAGTLLASHMENYFFDYDLKREMLR

Cpn_1049 1202586 1203914

lysC-Aspartokinase III
EGNVSKIIVYKFGGTSLAENICLVCDIICKDKPSFVWSAIAGVTDLLVDFCSSLRER
EEVLARKIEGKHIEVKNLAIPFVSTWTSRLPYLQHEISDLDFARILSGEDISASLV
RAVCTRIGWDLGLFEARSVILTDSSYRASPNLDLMAHWHQLELNQPSYIIQGFISGN
LGUTVLGRGSDYSATLIAELARATEVRIYTDVNGIYTDMPKVISDAQRIPELSEFMQ
NLASFGAKVLYPMLFPCMRAGIPFVTFSTDFEKGQTVWYAVDKSVSEYPRIKALSLSQ
YQSFCSVDYTVLGGCGLEELIGLLESHGIDPELMAQNNVGVFVMDDDIISQEAQEHVLD
VLSLSSTVRLLHSHVALITMIGDNLSPSKPVVSTITEKLRGQGVFCQSSMALSFFVAS
ELAEIGIEELHNDYVQKAIAT

Cpn_1050 1203884 1204798

dapA-Dihydrodipicolinate Synthase
LCKTKSVSRHVGRIHMLLTATVTPFFNGTIDFASLERLLSFQDAVGNVVLGSGTGGL
SLTKKEKQALICFACDLQKLVPLVCTSGTLLEEVLDWIHFCNDLPIISGLMTPTIYTKP
KLCCQILWFEAVLNAAKHAPATLYNIPSAARATPLYDVLKAAHHPFLGIDKDSGGSVEEF
OSYKSIAPHIQLYCGDDVFWSEMAACGAGLISVLNNAWPEAEYVLNPOQDYSLSWM
ETCRVYVITNPIGIKALAYKAITFTRKQTLPLSIEDFLENVSPAVEGMLAWPKLRTS
VFSYS

Cpn_1051 1204956 1205270

No robust homolog present in Genbank/EMBL as of 11/7/98
FFMTPKSIQGLHLIKTIDPVKISVPTTKSSFFRQSLRFLFLFWMFLYCRISIRPHCV
HIATFICQRLILTLTFLSMICLHFTLWLCEDKPIIRKNK

Cpn_1052 1205402 1206169

No robust homolog present in Genbank/EMBL as of 11/7/98
FFIQMKKYNREKIKALRICGTCITVERNIFLSCYDKIFYLGLYVFMFNSICRCR
GCGFRIGKLEVEYTKVKIKDEIPGELNDPVKVAESFPYRHALESLSGQSGICNLCA
LGNFLDKMLRNKIKHWTITFTRKQTNADGSEPPFYACVILAKRLKLAGSYEL
QVACGLQGLKQVDSHIRTATCGILVIRKMYTRLPCTPYIWKARPLFPFRLTSD
VRRDILKKFRLHECKD

Cpn_1053 1206198 1206701

No robust homolog present in Genbank/EMBL as of 11/7/98
KKKGLIAPKLIHNLVLTUAAZVACILTCLNIPYKAPAEHVLVSKRPIEGEP
SKLAPYNDPYIVSGPILGLEYETDGRIDHNDIMVAPKGAULRETRAKEITPG
SKDYAPFELTARELMEKELAMTAVSEVIGPCTYQCTYKTKNIKROYRIHJNTPG
ELGVKAPF

CpN_1054 1207010 1209466
No robust homolog present in Genbank/EMBL as of 11/7/98
SRMHHRFIMQVLLSPQLPPPPQHSVCSSISGSKRLVLAITFLVGLMLLLISGALFLLTGI
PGLSAAISFGLGIGLSALGGLVLMISGLLCILLVREIPTVRPEEIPGVSLAPSEEPALQA
AQKTLAQPKELDQDLDIQEVFACLRRLKDSKYESRSFLNDKAKELRVDFVVDLTSE
IFELQIVIAQEGWDLNFIINGGSSLMNTAESLSDLFHVSKRLGLVPSGVRGCEGLKSSA
KFTVAPIMSLHCEITHKVAFAFDNRYAMAEKAFKALGAEESVYSLTQSYRDKFLSE
KFTVAPIMSLHCEITHKVAFAFDNRYAMAEKAFKALGAEESVYSLTQSYRDKFLSE
YVADWYDQEFQKAGERLEKHLALYPEVSVSIRENKIQTNRNLEKAYEAIEENYRCVRE
QEDYWKKEEKRAEFERGRNKLSPPELESSELEQDFHGLKNFSEKLMELBGLHLLKQKEA
TAEVYNKILSDAESRLIEIVFDVKEMPCRIEIEKTLRMAELPLLPTKKAFAKACVQYNS
CAEMLEKVPYCKESLAYVTSEKRLVSLDEDLRRAYTECKRFGQDGLSEVRAEQL
REIRIQEFETQGLDLVEKELLVSSRLNTECDVSGVKEAPPGKFFAQYDEIYRVV
QSRWMTSERLREGVQACNKLKAGLSEEDKVLKEEYVLYERKKNKELVGTGKIVAT
QQRVAAFESIEVPEIPEAPEKPSLLDKARSLFTREDHT

CpN_1055 1209583 1210521
No robust homolog present in Genbank/EMBL as of 11/7/98
CKYLYHNSPPPPHDSVGAFFCLSKFRVLAITFLVGLVFLISGALFLLTGISLSAITS
FGLGIGLSALGGLVLSGLLCKLAKREVPVTRPEEIPGVSVAPSEEPALQATQKTLAQ
PKELDQDLDRIYQEVVCLGKLKDLRCEQDGLLDAKEKQVDFVWKDMHTEFVQLQIM
DQEGWYKLIQEMRDIGSTLFMSQVSLFKLWELGYLPSGVRGERLKKARSVDQRM
RRICDTKRVAMTDFRNAYGVAKTAFKAFGALETQVYKSMTESYREAFCEYKTKILRDE
EKILRICYLELRR

CpN_1056 1210482 1211228
No robust homolog present in Genbank/EMBL as of 11/7/98
GEDIKDMLSRVEEIDMLRVIELPLLPKQALEKAFVQYNSYKAKLTKEVPCFRESPAYI
TSEERLQSLDQTLERAYKQYKQFQPSRLESSEVSGREHLREQVKQFETQGLDLKEEL
IFVSDVLFKRMVSLVSTVHVPFMEFYEFELHRLRLRAQMANAEIYSKVRKAFPEML
KETLEKAKAPREEEYWLCEERKSKERLILNKIEAAQQRVLDLEPPPIKETGQKQKKE
YSFFIRLKS

CpN_1057 1211467 1213596
CT356 hypothetical protein

IIHFYFNFAMPEPLYTNKLTETKSPYLLLYAHTPVNMWYPMGAEAFHIAIENKPVFLSI
GCKHSRWQCVMLQESYTNPEIAAMLENYFVNKVDKEELPYVAKLYGDLAQMLAVSGDHQ
ETVSWPLNVLFPDLVPFFSVNYLNGEGLGLPSFPQIIDLKHFMEWDAEEREALEVMQAM
KVEIASFLEGCVKREILDESILKRTYVAAALYQDIDPHYGGVKAFFKRLPGLLLQFFLYS
LEYQESRGLFFVDRSLSMVALGGVRDHHGGGVYSYITDDKWLIPAFKRLIDNALMALNY
LEWACIGKEEYRGIGKILSYILSELYSPEVGAFFSSEQAENMGAGGQNFYTSVEEIS
NAGEADAEICDYVYISREGFFNGRNILHPIVHREIEELSEKYSRSEIAEDIVDRSDI
LKGTRARSHRSKDDLSLTFNNGMIITFYAYAGRLGVEVEYIEIGKCGEFVNSLYKHH
ELKRWREGEAKYRASLEDYDALILGLVLALESQCGSFWLSFAEELMQEVLVSFRSEEGG
FYSVDGRDSTLLIKQSPDSGETISGNALICQCLLSHLITEKHYLYTAEEDILQIAQAC
AHTKFSGLGLLIASQNYSRKHVKVLIAGLDQEDRSPLKCLSGFLPYLSLIMWTOEN
QEHLETVLPEYHCLIPKGDCTATTIYVLEVDQCKRFKDLLELFRYILSL

CpN_1058 1213742 1214836
CT355 hypothetical protein

EVNRYQTLRGIVLVSTGCIILGHHGGYAAEVPVTSYENLLESKEQDPSGLAIHDIRL
FVYDEENVVLTADVHKLNLFLYNSYPHLIDSPARSQYTTAMVPLVESIDEFLMVD
AKAKRIATDPTAVQNEIEEMFGRLDPLVAFHEMSPNDIFNVNDRTLTAQRVMGMVRSK
VMLKVTGPKIREYRKLEEAASRVVWKYRVLTIKANTEASQIADKVRALNEAKTWD
KDRLTALVLSQGGQLVCSEEFSSRENSLSQSHQKQDLIGYKELCGLPKAKHSGYKLYM
LLDKTSGSIEPLDMESKIKQHLFALEAESVEKQYKDLRLRKYGYDASMIKLLSEEA
LPSEEL

CpN_1059 1214848 1215678
KpA-Dimethyladenosine Transferase

VTRSSPAQLSRFLSEIQNKPKKLSQNFVLDQNVKIVATSEVIPQDWLEIGPFGAL
TEELIAGAQVIAIEKQMPFAPSLEELPLRIEIDACKYPLDQLEYKTLGKGRVANLP
YHITPLLLTKLEADPQKVTYVYMQDEARRIVAQPGGRDYGSLTIFLOFFADHYAF
VYSSCFYKPKQVQSAIVHMKVKTLPDSDEIIPVFTLTRTAFQRRKVLANKLGLPY
KEQVEQALKELGLLLNVRPEVLSNDYLLALFKHQAG

CpN_1060 1217694 1215727
dxs/tkt-Transketolase

YKRPLYIHITKMTSSSCFLDLLSPADLKKLSISQLPGLAEIYRILISVLSQTCGHL
SSNLGIVELTIALHYVSSPKDFIFDVGHQTYPHKLLTCRNNCGFDHNDNGLSGFTN
PTESDHLDFSGHAGTALSALGMAQITPLESRTHVPIPLGDAAFSCGTLLEALNNISTD
LSKPVVILNDNMSISKNGVAMSRIFSRNLHHPATNKLTKQVEKLAGIPRYGDSLAKHS
RLSQCVKMLFCPTPLFEQFGLAIVGPIDGHNVKKLIPILQSVRNLPPIILVHVTTKG
GLDQAQNNPAKYHGVANFNKRESAKHLPAIKPKPSFPDIPGQTLCLGEVSSRLHVTP
AMSIGRLEGFKQKPERFDFVGIAGHAVTFSGIAKAGNPVICSTYSTFLHRLDNVP
HDVCMQDLPIVIAIDRAGLAYGDRSHGIDYMSFLRAMPMIICQPRSQVVFQQLLYSS
LHWSSPSAIRYPNIPAPGDPDLTGPDLNFRSPGNAETLSQGEDVLIILGTLCTALSIK
HQLLAYGISATVVDPIKFPDNDLPSLLMSHVKVITIEHSYRGLASEFNNAVTFN
PKVDILNFAIPDTFLSHSGKALTKSILDESSMTNRLTHFNFRSKQTVGDVVR

CpN_1061 1217932 1217666
CT330 hypothetical protein

FGSLMVEIHHKDPGLKFLALQOSETLNSLSDIVATYEAFFSLIYEGNLKALRKDQCLY
LLSVNSKQELLKSPSGDPIVQTFPIHPHH

CpN_1062 1219835 1218159
xseA-ExoXyribonuclease VII

ROFPMVLEPPOAVASLTERIKTLLESNFQIIVKCELSMVSQKQXHLFGIKDSQAFLN
IAFFHEKQYKDYKPKQDAVIHCKLAVYAPRQGIIVAHALVYAGECDLLQKFEETR
KIAEAGYFATEKKKPLPFAQPGQIGVITPTGAVIQDLRLVLRARNRYILVYPTVVGON
IAHIEIKAIEVMAENADVLITAROGGSIEDLAFNEELLVKAHASTIPTVSAGHVE
TDYTLAFASDVAPTPPAAAEIVCKSGSEEQVVFQYRLHLSHJRLLTSSKKQQLPW
HPIIDRAEFTTAAQQLDSEIETAIQKUVQCKITHESKORYDNIIRWLGQDLVCPMTCLQS
IKMLLQALSHKALQLQVRQIKQLKGLTYPQQAQCKLWRQQLDLCRLPHYOKE
RYPHIKITPLKHAHNVLEQLRLHVKQLLELGLPCLPCECLNQLQKIAVANVETLATIL
KRYENIARYZALKQLHLINPKNVLRGYAMLFDFHENCAMLEVDSDQENARVIOLO
IGEAITLVNTIEICKLKG

CpN_1063 1219900 1220712

tpi3-Triosephosphate isomerase
FCRESMRKIFRENKRMTRFENWKMHKTIQEAKEVCTLAJLQDQEPLOCTIGIA
GPFTGLRAIHEVINTTGAFLLWCAQNVHPELSGAFYGEISLPMLEKV7VEFVLVGHSEER
HIFGSEDAFIAKVKVSAQAGLVVLCVGESLEVRREGKAQHV1KKQLLLGECMDNGSE
FLIAYEVPVIAIGTKVAAEASDVQDTHMFCREVVAERFSEATAEISILYGGSVKVDNAQR
FQCSDVDGLLVGGASLEGGQSFVAVKNNFN

CpN_1064 1220716 1220895
def-Polypeptide Deformylase

IQVVLVRDFTTELCOAHVOTMIRLEYYGSPILRKKSSPIAEITDEIRNLVSDMCDTMEA
HRCVGLAAVGVKNVSLFVMCVREDEDELIFSESPRVFINPVLSDPSETPIIGKEGGL
SIPGLRGVEFRQKITVYAMDNGKIFTEHLEGTARIIMHETHDLNGVLYDLMSEPKD
PKKFKASLEKIRRYNTHLSKEELVS

CpN_1065 1221140 1220928
No robust homolog present in Genbank/EMBL as of 11/7/98

RHRLGRHRTSDPCFLYFSIPEESLPDSCRLNMQPKHLEPLSILLKKPIIDLYKITSI
YEKAIFNTGLP

CpN_1066 1221132 1221488
No robust homolog present in Genbank/EMBL as of 11/7/98

SMNLNIGMTVLYAFILFLICVLICGLILVQESKSMGLSGSGVDSGDSVFGVSTP
DILKVVTSWCAVAFICIGLLLSFTNLGLKGLDAKEFLPAEESDQASSSESVEADES

CpN_1067 1221675 1222292
def-Polypeptide Deformylase

IQVVLVRDFTTELCOAHVOTMIRLEYYGSPILRKKSSPIAEITDEIRNLVSDMCDTMEA
HRCVGLAAVGVKNVSLFVMCVREDEDELIFSESPRVFINPVLSDPSETPIIGKEGGL
SIPGLRGVEFRQKITVYAMDNGKIFTEHLEGTARIIMHETHDLNGVLYDLMSEPKD
PKKFKASLEKIRRYNTHLSKEELVS

CpN_1068 1223267 1222365
rnhB-Ribonuclease HII

MSCMPPFFVYVLTTSAGNLLRDLQKEKNFISQPNVTFQARSNTVCTLYSPKLVIOG
KCGSEFTEFFLEPEILHTTHARVEQDLRPLRGVDESGKDFGFLPLCAAVYASNAEILK
KLYENKVDQSKNLDKTIASLARIIRSLCVDVILYPEKYNELYGKFNLTLLAWAHA
TVINNLAPKPGDVFAISDQFAASEYTLKALQKKTETITLIQKPAEQDVVAAASILA
RDFVQSIQKLEEGYQVQLPKAGFNVKAGREIAKQKELLAKISKTHFKTFDEICSG
K

CpN_1069 1223507 1223941
yfgA-HTH Transcriptional Regulator

VIMQHEKILLHLEIFRSRESQSLSKDVEAATSIRYSCLEATEEQGLGLKISPVYA
QFIKKYATYGLDGLSILQEPHYVMKIFKEFSDHNMMLLDESQGRNSPERAIHSWS
NLWAGLITIGIMVWMLGSLFSIF

CpN_1070 1225523 1224144
No robust homolog present in Genbank/EMBL as of 11/7/98

RRSLMTFPCGNCNYYRETPPNPGGEDIPLQEGGSGSGGGRVITQPGTQGRMGISL
GSDNVLGMVQAGSLNLLNLDARMORLGHYCYRTGTWPCREHCPGLQWIGGCCACCL
ETVDDPNPSAQFLOQLIQYGPICVGMFQQLPHCTQKIEQGEPLGDKQEVENGCKL
HRELLKAAQPRCMGESLVKLLQNNQNGEDMQTPPWSLILQAVSEGLASFVTSNDNP
WLOPQEQPCPPPTDEQLQAGVGAPOAKKHPAQECRVCTKLNFTLLQKLSRLVE
LSLESGVPLGQAAQIVDLIKKSLKRLVADSLATFLGPGIGLSLESQVFEVLVLCIL
SKGVLPLDPLHPEQTVLDPVQGPQWQRLRLVLTTRTANNIRQTOGEAPROAPPPDP
WDDEIERDGIIVTGGGFGIPQCCLRCWRKLPTEKRRNWL

CpN_1071 1227336 1225885
No robust homolog present in Genbank/EMBL as of 11/7/98

KGTMVCPNNSWFRMCGNFNCWVVTTEETROSASDISEAGSSGGAAPITPTQTKI
TKVEKRVQNTAQGDSTIHIQEAELVDSILSHRRTOQCTEYCYDSYATCGCQRCGSF
GRLICGTYKACCLDREDNQVAGLVHECEQTHGPIAVALAATMGLNLMELVEKMTILSEE
QKNEFRQHCSEAKTQLYGTMQSLQNFLEGVNSIRERGLDSDLVQAVLSFIATRSWEKT
IESEASQTSASNSTRIACYPILNTSPLTSLSCSGSDARRPSVGAEPQYVAKKYND
NGMARLGKIQVNLKTDGFSALGPFGLLVKMLNSFLLSASQTSLSLKHTGGEICYTC
PNFRDVLVLLMLAIGYCPANTDETSTVVDHMDIPMTIFYRLOYSYRTGKTSAFLKKK
PSLVQESLDCPTPAESVPLMSSLEEDENEDDEDGNLAYQQRILECSGHLQTLFLGK
INKE

CpN_1072 1227924 1228835
No robust homolog present in Genbank/EMBL as of 11/7/98

KKDYILHANWCCWQMLKIQKRMVSVVTVGAIVGFNSADAAPKKKKIPIQILYSPT
KVSSYLKNEASTIFCVDDVRLQLQRYLGSFGWQETRRRLFKSLNOSYGNRLEGET
LAIDIFRNKECLESEIPEQMEALANSSALVLGSISSFGITGIPATLHLSRLNLSFQKRS
IASSFSLKIDSAPSDASVYKGVLFRCGETAIVDALSQLFAQLDLPKKIIFLGEDEPVV
QAVGSACIGWGMNPLGLVYPAQESLSYVHPYSTATELQEAQGLQVISDEVAQLTNAL
PKMN

CpN_1073 1229011 1229832
Predicted OMP (CT371)

MRVYLFMLALCLYRAAPLEAVVITDAQAVLKFAREKTLVCFNIEDTVFPKMQVGOS
AWLYNRELDKTLSEEQAREQAFLEWMOISFLVDYELVLANRLV/LGSLKRSVVLGI
SQRPVHLIKNTLRILRSFNIDFTSCPAICEDCWLSHPTKDTTDOAMAIENKILPFGVSLK
NQCPMDAALEVLVLSGSISSPSSQIIYVDQDAERLSIGAFCKANKAIYFICMLYTPAKORVE
SYNPKLTAIQWSQIRKNSDEYESSLSYVSKS

RNA SECTION

tmRNA 13849 138074

Ribonuclease P RNA 697432 697649

16S rRNA 1009564 1002112

23S rRNA 1002415 1005278

5S rRNA 1005293 1005509

